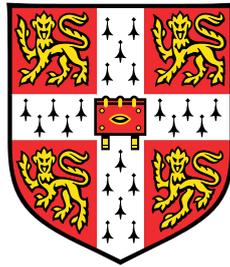


# Reconstructing the Sky with CMB Lensing methods



**Omar Darwish**

Supervisor: Dr. Blake Sherwin

Department of Applied Mathematics and Theoretical Physics  
University of Cambridge

This dissertation is submitted for the degree of  
*Doctor of Philosophy*

To my family.

## Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. Chapter 3 contains my original work published in [72], and was carried out in collaboration with Blake Sherwin, Mathew Madhavacheril, and the rest of the ACT collaboration. In chapter 3 I produced most of the plots and code, with code and inputs based on earlier ones from Blake Sherwin, Mathew Madhavacheril and the ACT collaboration. Multi-frequency cleaned CMB maps were created by Mathew Madhavacheril. The text of the chapter was written by myself, Blake Sherwin and Mathew Madhavacheril. Chapter 4 is my original work carried out in collaboration with Blake Sherwin, Noah Sailer, Emmanuel Schaan, and Simone Ferraro. I carried out all the calculations, wrote the code and produced all the plots. Emmanuel Schaan and Simone Ferraro provided the foregrounds and CMB lensing convergence cutouts from the simulations of [244]. The chapter was mainly written by me, with editing and comments by Blake Sherwin, Noah Sailer, Emmanuel Schaan, and Simone Ferraro. Chapter 5 is based on my work in a collaboration with Simon Foreman, Muntazir Abidi, Tobias Baldauf, Blake Sherwin and Daan Meerburg [71]. I performed, starting from previous work done by Simon Foreman, all the main calculations and numerical details. I wrote the core computational code to obtain most of the results of the chapter (with a couple of additions by Simon Foreman to, in particular, allow for a wedge for 21-cm forecasts). Simon Foreman used the code to produce forecasting plots based on fiducial experiments. The simulations were created and used by Tobias Baldauf and Muntazir Abidi, and the results of the simulations section are by Muntazir

Abidi. The writing of the chapter was a collaborative endeavour (in particular, I started the first draft, then I focused on a few sections, and checked the whole text). Finally, I wrote and obtained the results for most of the appendices, apart from the ones on reconstruction from the bispectrum by Blake Sherwin, 21-cm reconstruction by Simon Foreman, and the the appendix on the bispectrum and trispectrum noise was written by me and edited by Tobias Baldauf.

Omar Darwish  
November 2021

## Abstract

Current and future high significance CMB lensing-galaxy cross-correlations will soon precisely probe the growth of structure and provide powerful tests for parameter tensions, non-Gaussianity, neutrino mass, and modifications to  $\Lambda$ -CDM and gravity. However, such cross-correlations are sensitive to CMB foreground contamination, and the resulting biases represent an important challenge for cosmological analyses. In this thesis we describe new CMB lensing cross-correlation measurements and a novel multi-frequency cleaning-based technique to mitigate foreground effects in lensing maps. In particular, we present a tSZ cleaned cross-spectrum between Atacama Cosmology Telescope lensing and BOSS galaxies using a modified version of the standard quadratic estimator that retains most of the signal-to-noise.

To minimize lensing foreground biases, geometry-based methods that modify the form of the standard quadratic estimator have been developed as well. We discuss how to optimally combine these with multi-frequency cleaned data to mitigate the impact of foreground contamination while maintaining maximal significance. For a Simons Observatory-like experiment, we find that for the CMB lensing auto-spectrum it is possible to reduce the foreground induced bias by a large factor, compared to the standard quadratic estimator, at a modest noise cost. We find a similar result for a cross-correlation with an LSST-like sample, with a large reduction in bias at small noise cost.

Finally, we turn to the three-dimensional large scale structure of the Universe probed by galaxy surveys. We show how we can reconstruct the large-scale density field from galaxy modes using CMB lensing-inspired methods and use this reconstruction to improve constraints on local primordial non-Gaussianity.

## Acknowledgements

I am very grateful for the opportunity of meeting and interacting with several great people during my journey here.

First of all, I want to thank my supervisor Blake Sherwin. He is a great scientist, and a fine human being. I am very fortunate working with him. He is a fantastic supervisor. Interacting with him I have improved as a critical thinker, a presenter, and to listen more and talk less. Thank you for pushing me when necessary, giving me lots of freedom, sharing your ideas, and your passion in cosmology. I have learnt a lot from you.

I am fortunate to have worked with some other cosmologists. Thanks to Mathew Madhavacheril, that helped me with his code in the beginning of my journey here in Cambridge, and mentored me several times. Thanks to the ACTPol lensing group. Thanks to Tobias Baldauf, from which I learnt how to be more careful as a scientist. I also want to thank my other collaborators, Simon Foreman, Muntazir Abidi, Daniel Meerburg, Noah Sailer, Emmanuel Schaan, and Simone Ferraro. It was a pleasure sharing fun, and interesting discussions with you all.

Talking about discussions, I shared few offices with my colleague Theodore Steele, a source of weekly doses of entertaining conversations. No need to check the internet!

I was fortunate enough to do two internships related to machine learning and data science, meeting interesting people. I had the opportunity to work on “real” world projects. I want to thank James Fergusson for helping me finding these, and making sure I have a nice experience.

I am grateful to Churchill College for its hospitality, STFC, CDT seminars, and DAMTP, a great place where to learn, do research, and have lots of free meals.

I also thank, on behalf of my collaborators too, Shirley Ho, Alex Krolewski, Will Percival and Anthony Pullen for useful discussions regarding the BOSS survey, Colin Hill

---

for useful feedbacks on the CMB lensing and CMB foregrounds, Emanuele Castorina, Azadeh Moradinezhad Dizgah, Simone Ferraro, Mat Madhavacheril, Moritz Münchmeyer, Will Percival, Matias Zaldarriaga, and Hong-Ming Zhu for useful conversations with my collaborators. And I also thank Emanuele Castorina, Azadeh Moradinezhad Dizgah, Emmanuel Schaan and Marcel Schmittfull for thoughtful comments on a draft of chapter 5.

Throughout my work I extensively made use of standard python packages (numpy, scipy, matplotlib [132]), the LensQuest code,<sup>1</sup> symLens,<sup>2</sup> pixell,<sup>3</sup> mystic,<sup>4</sup> healpy [294], HEALPix [101], Astropy [20, 216],<sup>5</sup> and CAMB [154] for calculating theory spectra. I have also used the National Energy Research Scientific Computing Center (NERSC) and Niagara supercomputers at the SciNet HPC Consortium.

I have always created problems as a student. Still, during my academic career several professors helped me, from Scuola Galileiana's heads, Paolo Ciatti, Michele Liguori, Sabino Matarrese, and finally Blake Sherwin, whose patience has been tested throughout these years. Without you, I wouldn't be writing this.

Being a student can be very difficult and stressful, especially in a foreign country. I want to thank from my heart Filippo Bernardi, Jakub Kaczmarczyk and João Alves. You made my first years much lighter, and we had lots of fun together. Special mention goes to Enrico Andriolo, with whom I shared deep conversations, and good experiences both in Padua, and then here in the UK.

I also want to thank The Chess Club for the best strategies and the dinners, Santiago Dubov for language and life discussions, and others, in particular the super fun Riccardo Gonzo. Thanks too to all the other people I met here in Cambridge!

Finally, mention of honor goes to Ustaz Hassan. He taught me for free, taking from his time, how to read and memorize the Qur'an by heart. It has been a difficult journey, but very rewarding. Even if we live in an era where we have other sources for interpreting the world, it still a great source of wisdom, and it has helped me memorize better topics by understanding.

The last but not the least comes my family, to whom this thesis is dedicated. My two brothers, Ali and Abdelrahman. And my parents, Nousa and Abdalla, who have worked hard and made their best to give their children a good life, and a better opportunity to navigate society, while away from their native home and their own families. Thank you so much for the troubles you had to overcome.

---

<sup>1</sup><https://github.com/EmmanuelSchaan/LensQuEst>

<sup>2</sup><https://github.com/simonsobs/symLens>

<sup>3</sup><https://github.com/simonsobs/pixell>

<sup>4</sup><https://github.com/uqfoundation/mystic>

<sup>5</sup><http://www.astropy.org>

# Table of contents

List of figures	xii
<b>1 Introduction</b>	<b>1</b>
<b>2 Review of Cosmological Perturbations</b>	<b>4</b>
2.1 The Expanding Universe . . . . .	4
2.1.1 The FLRW Metric . . . . .	4
2.2 How do we describe cosmological observables? . . . . .	10
2.2.1 Fourier space . . . . .	12
2.2.2 Cosmological perturbations: gauge issue . . . . .	13
2.3 How perturbations may arise . . . . .	14
2.3.1 Seeding perturbations with inflation . . . . .	15
2.4 How the Universe looks today . . . . .	21
2.4.1 CMB fluctuations . . . . .	22
2.4.2 Boltzmann equation . . . . .	25
2.5 Secondary Anisotropies of the CMB . . . . .	38
2.5.1 Foregrounds . . . . .	39
2.5.2 A component separation method . . . . .	42
2.6 CMB Weak Lensing . . . . .	44
2.6.1 The CMB lensing potential . . . . .	45
2.6.2 The lensed CMB . . . . .	48
2.7 Dark Matter and Biasing . . . . .	53
2.7.1 Late-time matter field . . . . .	53
2.7.2 Biasing . . . . .	59
2.8 Reconstructing modes with quadratic estimators . . . . .	65

---

2.8.1	CMB lensing reconstruction . . . . .	66
<b>3</b>	<b>The Atacama Cosmology Telescope: A CMB lensing mass map over 2100 square degrees of sky and its cross-correlation with BOSS-CMASS galaxies</b>	<b>70</b>
3.1	Introduction . . . . .	70
3.2	Theoretical Background . . . . .	72
3.3	Lensing Maps from ACT Data Alone . . . . .	73
3.3.1	CMB maps for lensing analysis . . . . .	73
3.3.2	Lensing reconstruction and validation . . . . .	75
3.3.3	Visualization of the maps and their correlation with large-scale structure . . . . .	78
3.4	Foreground-mitigated lensing maps with new cleaning methods . . .	80
3.4.1	A new tSZ-free estimator . . . . .	81
3.4.2	Application to data . . . . .	82
3.5	Galaxy cross-correlation measurement . . . . .	84
3.5.1	The CMASS galaxy map . . . . .	84
3.5.2	Extracting power spectra and obtaining the covariance matrix	86
3.5.3	Galaxy cross-correlation: results . . . . .	87
3.6	Systematics and Validation of the cross-correlation measurement . . . . .	90
3.7	Discussion . . . . .	95
<b>4</b>	<b>Optimizing foreground mitigation for CMB lensing with combined multifrequency and geometric methods</b>	<b>97</b>
4.1	Introduction . . . . .	98
4.2	CMB Lensing foreground biases for estimators . . . . .	99
4.2.1	Quadratic CMB Lensing Estimators . . . . .	99
4.2.2	Origin of foreground biases . . . . .	99
4.2.3	Current foreground mitigation methods and estimators . . . .	100
4.3	Optimal combination of CMB lensing estimators: formalism . . . . .	102
4.3.1	Amplitude shift on the CMB lensing power . . . . .	102
4.3.2	Amplitude shift on CMB lensing power from composing geometrical and multi-frequency methods . . . . .	104
4.3.3	Amplitude shift on CMB lensing power for combined estimators	106
4.4	Finding the optimal combination . . . . .	108
4.5	Optimisation in practice . . . . .	109
4.5.1	Calculating foreground CMB lensing biases . . . . .	109

---

4.5.2	Results . . . . .	110
4.6	Conclusions . . . . .	120
<b>5</b>	<b>Density reconstruction from biased tracers and its application to primordial non-Gaussianity</b>	<b>125</b>
5.1	Introduction . . . . .	126
5.2	Density reconstruction . . . . .	132
5.2.1	Quadratic estimator: general case . . . . .	132
5.2.2	Non-Gaussianity and bias expansion . . . . .	136
5.2.3	Reconstruction noise and contamination . . . . .	141
5.3	Simulations . . . . .	143
5.3.1	Generation of quadratic estimators . . . . .	144
5.3.2	Cross-correlation of quadratic estimators with the initial linear field . . . . .	145
5.3.3	Auto- and cross-correlations of quadratic estimators . . . . .	147
5.3.4	Visualization of reconstructed field . . . . .	149
5.4	Forecasts . . . . .	151
5.4.1	Fisher matrix setup . . . . .	151
5.4.2	Analytical derivation of cosmic variance cancellation . . . . .	153
5.4.3	Assumptions and experimental configurations . . . . .	155
5.4.4	Expected precision on reconstructed modes . . . . .	159
5.4.5	Results: constraints on non-Gaussianity . . . . .	161
5.5	Discussion . . . . .	168
5.6	Conclusions . . . . .	170
<b>6</b>	<b>Conclusions and Outlook</b>	<b>173</b>
6.1	Conclusions . . . . .	173
6.2	Outlook . . . . .	174
<b>Appendix A</b>	<b>Appendices of Chapter 3</b>	<b>176</b>
A.1	CMB map pre-processing for lensing reconstruction . . . . .	176
A.2	Noise properties of the symmetric foreground-cleaned estimator . . . . .	178
A.3	Data Availability . . . . .	181
<b>Appendix B</b>	<b>Appendices of Chapter 4</b>	<b>182</b>
B.1	CMB lensing biases for a combination of estimators . . . . .	182
B.2	Composition of CMB lensing estimators . . . . .	189
B.2.1	The symmetric estimator . . . . .	189
B.2.2	Bias Hardening . . . . .	191

---

B.3	Optimization explorations . . . . .	192
B.4	Lensing validation . . . . .	194
B.5	Optimization details . . . . .	196
B.5.1	Choosing among different loss functions . . . . .	196
B.5.2	Optimization algorithm . . . . .	199
B.5.3	Estimators over which to optimize . . . . .	199
B.6	Understanding the weights . . . . .	202
B.7	CMB lensing biases from quadratic estimators . . . . .	204
B.7.1	Biases from simulations . . . . .	204
<b>Appendix C Appendices of Chapter 5</b>		<b>206</b>
C.1	Derivation of density reconstruction from the bispectrum . . . . .	206
C.2	Noise expressions for quadratic estimator . . . . .	208
C.2.1	Noise for the auto correlation of the reconstructed field . . . . .	209
C.2.2	Noise for the cross correlation of the reconstructed field with the tracer field . . . . .	214
C.3	Bias-hardening . . . . .	215
C.3.1	General derivation . . . . .	215
C.3.2	Application to long-mode reconstruction . . . . .	218
C.4	Contamination of quadratic estimator by $f_{\text{NL}}$ terms: analytical expressions	220
C.4.1	Expansion of basic quantities . . . . .	220
C.4.2	Expansion of mode coupling expressions . . . . .	221
C.4.3	Writing the large scale contamination terms . . . . .	223
C.5	Foregrounds for 21 cm intensity mapping . . . . .	225
C.5.1	Implementation in forecasts . . . . .	225
C.5.2	PUMA forecasts with $K_{\parallel, \text{min}}$ . . . . .	228
<b>References</b>		<b>229</b>

## List of figures

2.1	A one-dimensional density field evolving in time. On the $y$ -axis, is the time slice, on the $x$ -axis, the position. We can see that the field is spatially constant in these coordinates. If we now choose another time slicing, represented by $\tilde{\eta}$ , we see that for the new time coordinate $\tilde{\eta}$ , the field is no longer constant. This figure is taken from [121]. . . . .	14
2.2	Temperature fluctuations from the cosmic microwave background (CMB) that were mapped by the Planck Satellite experiment. . . . .	22
2.3	CMB (lensed) band powers from recent ACTPol [63], Planck [8], and SPT [269] measurements marginalized over galactic and extragalactic foreground emission and Sunyaev-Zel'dovich effects. In grey, a theoretical curve from parameters of a best-fit of Planck in $\Lambda$ CDM. Image taken from LAMBDA <a href="https://lambda.gsfc.nasa.gov/graphics/">https://lambda.gsfc.nasa.gov/graphics/</a> . . . . .	36
2.4	CMB (lensed) power spectrum for a varying baryon density term, calculated from a best-fit cosmology given by Planck [8], using CAMB [154]. In black $\Omega_b = 0.022$ , red $\Omega_b = 2 \times 0.022$ , and in green Planck binned band powers as from LAMBDA <a href="https://lambda.gsfc.nasa.gov/graphics/tt_spectrum/TT_data_2020aug_csv_format.dat">https://lambda.gsfc.nasa.gov/graphics/tt_spectrum/TT_data_2020aug_csv_format.dat</a> . We can see the shift in acoustic peaks based on the baryon density term. . . . .	37
2.5	CMB (lensed) power spectrum for a varying curvature term, calculated from a best-fit cosmology given by Planck [8], using CAMB [154]. In black $\Omega_k = 0$ , red $\Omega_k = 0.1$ , and in green Planck binned band powers as from LAMBDA <a href="https://lambda.gsfc.nasa.gov/graphics/tt_spectrum/TT_data_2020aug_csv_format.dat">https://lambda.gsfc.nasa.gov/graphics/tt_spectrum/TT_data_2020aug_csv_format.dat</a> . We can see the shift in acoustic peaks based on the curvature term. . . . .	38

- 2.6 Comparison between Planck only, ACT+ Planck ,and ACT only for a single frequency (90 GHz) in a small,  $3 \times 3 \text{ deg}^2$  patch of the sky. We can see how the higher resolution of ACT allows for detection of small-scale features, such as clusters, compared to Planck, which by contrast captures better the large scales compared to ACT alone. The figure is taken from [193], where there are many other examples that emphasize the difference between Planck and ACT on small scales. . . . . 39
- 2.7 The effect of massive neutrinos on the CMB lensing potential as we increase  $\sum m_\nu$  from 0 meV to 150 meV. We fix other cosmological parameters in producing this plot, using CAMB [154]. . . . . 47
- 2.8 Top panel: CMB  $TT$  power spectrum for two geometrically degenerate models: in blue, one with a curved Universe and no cosmological constant ( $\Omega_\Lambda = 0, \Omega_m = 1.29$ ); and in red a flat  $\Lambda$ CDM Universe ( $\Omega_\Lambda = 0.73, \Omega_m = 0.27$ ). The black dots are the seven-year WMAP temperature power spectrum data [147]. We can see that the data do not significantly favor either model. Lower panel: we show CMB lensing deflection power spectra for the same two models. We can see that they are no longer degenerate, with the  $\Omega_\Lambda = 0$  universe producing a lensing power spectrum larger than that measured by ACT data. Figure reproduced from [253]. . . . . 49
- 2.9 The fractional difference between lensed and unlensed CMB temperature power spectra, with respect to unlensed CMB power spectrum. . . . . 51
- 52figure.caption.10
- 2.11 Example of matter power spectrum. Solid lines are for  $z = 0$ , dashed ones for  $z = 0.8$ . Black is the linearly evolved power spectrum, while red accounts for non-linear corrections in the power spectrum. We can see how clustering changes going from dashed,  $z = 0.8$ , to solid  $z = 0$ , and how small scales are affected by non-linearity, with an enhanced power spectrum with respect to the linear case. Figure taken from <https://camb.readthedocs.io/en/latest/CAMBdemo.html>, using spectra calculated from CAMB. . . . . 58
- 2.12 In red, the two-season ACTPol measured lensing power spectrum. In black, the theoretical best fit in  $\Lambda$ CDM CMB lensing power spectrum  $C_L^{\phi\phi}$ . Figure from [256]. . . . . 69

- 3.1 Verification of our lensing reconstruction pipeline for the tSZ-free lensing maps (shown for the D56 patch). We plot the average cross-spectrum of the reconstructed lensing maps with the input lensing simulations (blue dots), the average power spectrum of the input lensing simulations (red crosses) and a binned lensing power theory curve in black. (The BN patch gives quantitatively similar results.) The bottom panel shows the fractional difference of the input-reconstruction cross-correlation relative to the input lensing power. The ACT only simulations give residuals of similar magnitude. From the good agreement of the input-reconstruction cross-correlation with the input lensing power, we can see that the true lensing signal in the simulations is recovered within percent-level accuracy; we absorb only a small correction into a simulation-based re-normalization. . . . . 77
- 3.2 Map of the reconstructed lensing potential in the D56 region (upper panel) and the BN region (lower panel) after Wiener filtering, shown in greyscale. (The lensing maps shown are the tSZ-cleaned maps combining *Planck* and ACT, although the ACT-only lensing maps appear similar.) Overlaid, we also show contours of an identically filtered but completely independent cosmic infrared background map (*Planck* GNILC 545 GHz). Since the correlation between CMB lensing and the cosmic infrared background (CIB) is very high and since our CMB lensing map has high signal-to-noise ratio on large scales, the correspondence between the lensing potential and the CIB can be seen clearly. Parts of the CIB map contaminated by Galactic dust have been masked in the BN CIB contours for this visualization, using a mask derived from the *Planck* PR2 Commander thermal dust emission map. . . . . 79
- 3.3 Lensing reconstruction test, as for Figure (3.1), but now correlating with a simulated galaxy field instead of the input lensing convergence field. The shaded region shows the multipole range used for the cosmological analysis. The lower panel shows the fractional difference with respect to the theory curve. (The BN patch gives similar results.) . . . . . 85

- 3.4 The cross-correlation between CMASS galaxies and CMB lensing convergence reconstructions from ACT. The cross-correlation measurements in the D56 and BN patches are coadded to obtain these results. The red points show results using an ACT-only lensing map, the blue points show results using a lensing map that has been tSZ cleaned. The multipole values for different versions of lensing maps are slightly offset for visualisation purposes. See Figure (3.5) for comparison with a theory curve fit to the cleaned measurement. . . . . 88
- 3.5 This plot shows our main CMB lensing – BOSS galaxy cross-correlation measurement with the ACT+*Planck* tSZ free lensing maps (blue points). A *Planck*-cosmology (and fiducial galaxy bias  $b_{\text{fid}} = 2$ ) theory template, with a free amplitude fit to the data ( $A = b/b_{\text{fid}}$ ), is also indicated with a dashed line. The green dashed theory curve is fit only over a restricted analysis range (shaded region for scales  $100 < L < 1000$ ); the black solid curve is fit over the full  $L$  range shown in this plot. (The bandpowers are nearly independent, with the off-diagonal elements of the covariance matrix showing correlations of less than 13%). We find good consistency in both cases with the *Planck*-cosmology derived theory template. . . . . 89
- 3.6 A curl null test: verification that the extracted curl-lensing – galaxy cross-correlation, which should be negligibly small in the absence of systematic errors, is consistent with the null hypothesis. The results shown are for a combination of both D56 and BN patches. The  $\chi^2$  probability-to-exceed (PTE) for this null test is also shown in the legend. 91
- 3.7 A null test verifying that cross-correlating the lensing map on one field with the galaxy map on the other field (and combining both spectra) gives a signal consistent with zero. Red points show ACT-only results, blue points show tSZ-deprojected lensing results. The fact that the PTEs in both cases are consistent with zero signal supports the conclusion that our uncertainty calculations are correct. . . . . 91

- 3.8 An extragalactic foreground null-test for the cleaned maps. We show the difference between the cross-correlations of CMASS with the tSZ-deprojected lensing maps for the cases of  $l_{\text{CMB,max}} = 3000$  and  $l_{\text{CMB,max}} = 3350$ , where  $l_{\text{CMB,max}}$  is the maximum CMB multipole used in the lensing reconstruction. Since extragalactic foregrounds rise rapidly towards high  $l$ , a substantial foreground residual in the cross-correlation would cause a null-test failure. However, our null-test results shown here are consistent with zero contamination for both fields (blue: points for BN, orange: points for D56). . . . . 92
- 3.9 The same test as shown in Figure (3.8), but applied to the ACT-only maps which have not been foreground cleaned (blue: points for BN, orange: points for D56). The PTE for BN shows a (mild) failure of the null test, as is expected if foreground residuals are important. . . . . 93
- 4.1 The temperature power spectra of different multi-frequency combinations of frequency maps for a Simons Observatory-like experiment: in red is the standard internal linear combination (ILC); in blue is the ILC with CIB deprojection; in orange is the ILC with tSZ deprojection. In cyan we show the cross correlation between the tSZ-deprojected and CIB-deprojected combinations. In black is the lensed CMB theory curve. For the tSZ-deprojected curve in orange, the large power on large scales is due to atmospheric noise. . . . . 101
- 4.2 CMB lensing reconstruction noise curves for the standard quadratic estimator (QE), shear (SH), point source hardened estimator (PSH), and the profile hardened estimator with a tSZ profile (PH), all applied to a minimum variance temperature-only ILC map, as well as noise curves for the standard symmetric estimator with tSZ-deprojection (Symm) and two new estimators that we introduce later in the text: point source hardening on the symmetric estimator with tSZ-deprojection (SymmPSH), profile hardening on the symmetric estimator with CIB-deprojection (SymmPH), and finally the minimum variance estimator with both temperature and polarization (QE MV). For the temperature maps we take  $\ell_{\text{min}}, \ell_{\text{max}} = 30, 3500$  in this figure, while for polarization we assume  $\ell_{\text{min}} = 30, \ell_{\text{max,pol}} = 5000$ . In black we show the CMB lensing signal  $C_L^{\kappa\kappa}$ . In the bottom panel we show the total bias to the CMB lensing autospectrum (as a fraction of the signal power) for some of the estimators. The relative statistical error for the standard QE is shown in grey. Note that the standard QE is biased at most scales. . . . . 103

- 4.3 Values of measurement standard deviation  $\sigma$  and bias  $b$  for the CMB lensing power spectrum amplitude  $A$ , assuming only temperature ( $TT$ ) reconstruction is used in the lensing power spectrum measurement. Results are shown for the different estimators with different types of foreground mitigation that have been proposed so far; several values are assumed and plotted for the maximum CMB multipole used in the lensing reconstruction. In this and the following plots, [QE, SH, PSH, PH, Symm] indicate the standard quadratic estimator, shear estimator, point source hardened estimator, profile hardened estimator, and symmetric tSZ-deprojected estimator respectively. The colored bands represent regimes where the bias is greater than a certain fraction of the statistical noise. We can see clearly that a trade-off between bias and noise exists for the different estimators proposed thus far: estimators that aim to reduce biases unfortunately incur a significant noise penalty. We would like to be able to reduce both foreground bias and variance of lensing estimators simultaneously (obtaining new estimators that are somewhat closer to the lower left corner of this plot). 105
- 4.4 As for Figure 4.3, but with the addition of two new combined estimators: i) SymmPSH, which stands for point source hardening applied to a symmetric estimator with tSZ deprojected; ii) SymmPH, given by profile hardening applied to a symmetric estimator with CIB deprojected. We can see that, for some configurations, these new estimators allow us to obtain somewhat lower bias at the same noise in comparison with the existing estimators shown in Figure 4.3, although potential remains for further improvements. . . . . 106
- 4.5 As for Figures 4.3 and 4.4, but now including results from performing an optimized linear combination of individual estimators (in this plot we only consider lensing measurement from temperature). The solid black dots show the optimized points, with results shown for optimization with  $f_b = 0, 0.1, 1, 2, 4$  going from dots on the right to the left (recall that  $f_b$  parametrizes the importance of bias-squared relative to variance in the optimization). We can see that the black dots representing the linear combinations of estimators perform significantly better than single estimators; using these optimized linear combinations we are able to reach a regime with a negligible bias with respect to the noise at only a modest noise cost. . . . . 111

- 4.6 Composition of the optimized linear combination estimators shown in the solid black points in the previous figure. The pie charts show the fractional contribution made by each individual estimator in the linear combination, calculated as  $\int_L w(L)a_i(L)$ . In the pie charts, colors indicate the type of estimator and shading indicates the maximum CMB multipole used in the reconstruction. When assuming  $f_b = 0$ , i.e. optimizing for minimum variance alone, as expected the standard QE alone is dominant. For an optimization which heavily penalizes bias using  $f_b = 4$ , a combination of PH and SymmPH is the best combination; this combination only has a 0.5% bias, although the signal-to-noise falls below 70. We may notice that for high  $f_b$  PH with a high  $l_{\max}$  is chosen, and not with a low one that we might expect to give lower bias. The reason is that at the important scales for the calculation of the CMB lensing amplitude, the PH with higher  $l_{\max}$  is lower in bias compared to the ones with lower  $l_{\max}$ . This is because there are internal cancellations in the calculation of the bias, depending on the specific  $l_{\max}$  (and set of simulations). . . . . 112
- 4.7 More detailed illustration of the estimator linear combination for the case of temperature-lensing-only power spectrum optimization with  $f_b = 1$ . The top panel shows the signal-to-noise ratio squared per mode for the lensing reconstruction, the middle panel shows the Gaussian-smoothed absolute value of the total foreground bias for each estimator. On the bottom, the weights per mode are shown for each estimator; this illustrates the relative contribution of each estimator to the linear combination. It can be seen that, for  $f_b = 1$  where bias and variance are assigned equal importance, the optimization tries to compromise between bias and noise per mode. Most of the constraining power for the CMB lensing amplitude derives from  $L \leq 500$ . . . . . 113
- 4.8 As for Figure 4.7, but here showing results for an optimization run with  $f_b = 4$ , i.e. the bias is assigned approximately four times more importance in the optimization than the noise. It can be seen that, as expected, this optimization effectively selects the estimator at each  $L$  that has the lowest bias per mode. . . . . 114

- 4.9 Linear-combination optimization results for the lensing power spectrum as in Figure 4.6; however, in this plot, we consider not just temperature data but are also including polarization data. As before, the pie charts represent the contribution from each estimator, calculated as  $\int_{\bar{L}} w(L)a_i(L)$ . We see that when bias-removal is prioritized, at left, the optimization selects a combination of geometric and multi-frequency cleaning, namely Shear and SymmPSH. This is different than the result for the optimal temperature ( $TT$ ) only combination, due to the higher importance of the primary bias term when also including polarization data as explained in the text. . . . . 115
- 4.10 Illustration of the estimator linear combination for power spectrum optimization with  $f_b = 1$  – now assuming both polarization and temperature data are used in the reconstruction. The top panel shows the signal-to-noise ratio squared per mode for the lensing reconstruction. The middle panel shows the Gaussian-smoothed absolute value of the total foreground bias (equation (4.19)) for each estimator and the combination. In the bottom panel, the weights per mode are shown for each estimator. It can be seen that, for  $f_b = 1$  where bias and variance are assigned equal importance, the optimization tries to compromise between bias and noise per mode. . . . . 116
- 4.11 As for Figure 4.10, but now showing results for an  $f_b = 4$  optimisation where bias reduction is prioritized. As expected, generally, the estimator with lowest bias gets selected in the combination. . . . . 117
- 4.12 The primary contribution  $\langle \kappa Q_{TT+\text{pol}}[T_f, T_f] \rangle$  relative to the CMB lensing autospectrum signal for the sum of foreground components, for  $l_{\text{max}} = 3500$  for a few estimators. This type of plot becomes important when adding polarization data, as the estimators with the lowest primary bias contribution tend to become favoured in this case, in comparison to the  $TT$ -only case. . . . . 118
- 4.13 Optimization results for the cross-correlation of lensing from temperature-only ( $TT$ ) data with an LSST-like sample. The pie charts represent the contribution from each estimator, calculated as  $\int_{\bar{L}} w(L)a_i(L)$ . We see that for a high deprojection of the bias, at far left, the optimization selects a combination of geometric and multi-frequency cleaning, namely PH and SymmPSH. . . . . 120

- 4.14 Optimization results for the cross correlation of lensing from polarization and temperature data with an LSST-like sample. The pie charts represent the contribution from each estimator, calculated as  $\int_{\bar{L}} w(L)a_i(L)$ . As the primary cross-bias has no contributions from (foreground-free) polarization, we can see that for high deprojection of the bias, at far left, the results are very similar to the  $TT$  only cross-correlation case. 121
- 4.15 Can the (sometimes complex) optimal linear combinations of estimators be easily simplified? Black dots show the optimal points for the CMB lensing auto-spectrum shown in previous plots. The grey cross is the shift in this optimal point, if we choose to simplify the combination by using only the two estimators which contribute most. In the top panel we show the  $TT$ -only case, and in the bottom panel we also include polarization data. We can see that in general we have at most percent level shifts, going from black dot to grey cross, so that simplifying the estimator combinations works well. . . . . 123
- 4.16 As for Figure 4.15, but for cross-correlations with LSST galaxies. We again conclude that the combined estimators can be simplified without significantly degrading their performance. . . . . 124
- 5.1 Reconstruction noise power spectra for estimators that use each of the quadratic mode-couplings discussed in Sec. 5.2.2. We omit curves for the  $c_{01}$  and  $c_{02}$  estimators, which are greater than the upper limit of the plot. The G ("growth") estimator has the lowest noise by far. These curves are computed for a DESI-like survey, but the hierarchy between them is unchanged for the other surveys we consider. The signal to noise on reconstructed modes (not shown) is likewise much higher for the G estimator than for S or T, justifying our use of the G estimator for our main results. . . . . 141

- 5.2 *Left:* Contamination in the expectation value of G estimator, corresponding to separate multiplicative biases on the amplitude of a reconstructed mode, computed for a DESI-like survey. Blue solid line is the estimator growth bias shown for comparison. Dashed lines indicate negative values. Several of these curves inherit the  $k^{-2}$  scaling of the scale-dependent bias in  $\delta_g$  arising from nonzero  $f_{\text{NL}}$ , implying that reconstructed modes can be used to constrain  $f_{\text{NL}}$  in the same way. *Right:* Ratio of scale-dependent bias from  $f_{\text{NL}}$  (for a fiducial value of  $f_{\text{NL}} = 1$ ) to total bias for  $\delta_g$  (solid) and  $\delta_r$  (dot-dashed). Local primordial non-Gaussianity has roughly the same relative contribution to the bias of  $\delta_g$  or reconstructed modes. . . . . 142
- 5.3 Cross correlations of estimators  $\hat{\Delta}_\alpha$  corresponding to the growth, shift, and tidal mode-couplings with the linear density field  $\delta_1$ . We compare theory predictions (lines) with simulations (points) for three different smoothing scales,  $R = 20h^{-1}$  Mpc,  $R = 10h^{-1}$  Mpc and  $R = 4h^{-1}$  Mpc, corresponding to maximum wavenumbers  $k_{\text{max}} = 0.05h$  Mpc $^{-1}$ ,  $0.1h$  Mpc $^{-1}$ , and  $0.25h$  Mpc $^{-1}$  respectively. In this figure, we plot  $\langle \hat{\Delta}_\alpha \delta_1 \rangle / N_{\alpha\alpha}$  (in contrast to what is defined in Eq. (5.17), in simulations we define the estimators  $\hat{\Delta}_\alpha$  without a prefactor  $N_{\alpha\alpha}$ ). We find very good agreement for the growth estimator for all smoothing scales, and also reasonably good agreement for the other estimators. . . . . 145
- 5.4 Auto-correlations of the quadratic estimators  $\hat{\Delta}_\alpha$ , for the same smoothing scales shown in Fig. 5.3. The predictions for the growth estimator agree with simulations for all smoothing scales. However, for other estimators predictions agree with simulations for large smoothing scales but for the low smoothing scales, the predictions slightly disagree with simulation results as higher-order terms become more important. . . . 146
- 5.5 Comparison of the auto power spectrum of the growth estimator  $\hat{\Delta}_G$ , normalised by  $N_{GG}$  computed from theory (in red), with  $(\langle \hat{\Delta}_G \delta_1 \rangle)^2 / P_{\text{lin}}$  (in blue). We compare simulation results (points) with theory predictions (lines) for the same smoothing scales as Figs. 5.3 and 5.4. We again find excellent agreement between simulations and theory. In the bottom right panel, we plot the cross-correlation coefficient  $r_{G\delta_1}$  between the growth estimator and the linear density field for three smoothing scales. We see that  $r_{G\delta_1} > 0.9$  for  $R = 4h^{-1}$  Mpc which is why in the bottom-left panel,  $\langle \hat{\Delta}_G \hat{\Delta}_G \rangle$  is signal dominated. . . . . 148

- 5.6 2D slices of the 3D linear density field (left panel) and the growth estimator  $\hat{\Delta}_G$  (right panel). For the growth estimator we used  $R = 4h^{-1}$  Mpc smoothing which corresponds to  $k_{\max} = 0.25h \text{ Mpc}^{-1}$ . We apply an external smoothing of  $R = 20h^{-1}$  Mpc to both the linear and reconstructed fields. As expected, we find that the reconstruction reproduces many of the large-scale features in the linear density field. 150
- 5.7 Probability distribution functions (histograms) of the linear density field and the reconstructed field from the halo density field of mass bin I. As in Fig. 5.6, we use  $k_{\max} = 0.25h \text{ Mpc}^{-1}$  for the reconstruction and apply an external smoothing scale of  $R = 20h^{-1}$  Mpc to both the linear field and the reconstructed field. The PDFs of the reconstructed field are scaled to have the same variance as the linear field, and shifted to have mean 0. We find that the PDF of the reconstructed field is very close to Gaussian. Note that here we have applied a low- $k$  cutoff to the modes used for reconstruction of  $k_{\min} = 0.05h \text{ Mpc}^{-1}$  in order to match the approach in our forecast section below. . . . . 150
- 5.8 Expected errorbars on the reconstructed power spectrum  $P_{\text{rr}}$  for the surveys and redshift bins we consider (*blue*), along with cosmic-variance-limited errorbars (*orange*), computed for bandpowers with  $\Delta K = 0.002h \text{ Mpc}^{-1}$ . Downward arrows indicate errorbars whose lower limits fall outside of the  $y$  axis range. High-precision measurements of the power spectrum of reconstructed modes will be possible in several cases, even in the presence of a 21 cm foreground wedge for PUMA. . . . . 160
- 5.9 Forecasts for a DESI-like survey. *Left*: Signal and noise power spectra involved in the forecast. The galaxy auto spectrum is well above the shot noise, while the auto spectrum of reconstructed modes ( $P_{\text{rr}}$ ) is roughly an order of magnitude below both the reconstruction noise ( $N_{\text{GG}}$ ) in the quadratic estimator and the shot noise contribution to the estimator variance. *Center*: Expected constraints on  $f_{\text{NL}}$  when only  $\delta_{\text{g}}$  is used (*solid*), or when  $\delta_{\text{r}}$  is also used (*dotted*). We assume that  $\delta_{\text{g}}(\vec{K})$  cannot be directly measured for  $K < K_{\min}$ , and marginalize over the  $b_1$ ,  $b_2$ , and  $b_{s2}$  bias parameters. *Right*: Ratio of  $\delta_{\text{g}} + \delta_{\text{r}}$  and  $\delta_{\text{g}}$ -only cases from the center panel. We only notice an improvement for higher values of  $K_{\min}$ , corresponding to using  $\delta_{\text{r}}$  but not  $\delta_{\text{g}}$  at  $K < K_{\min}$ . . . . . 161

- 5.10 The analog of the right panel of Fig. 5.9, with a variety of (mostly artificial) modifications to the forecasts. There is no improvement in  $\sigma(f_{\text{NL}})$  when  $\delta_r$  is neglected at  $K < K_{\text{min}}$ , indicating that the inclusion of  $\delta_r$  at  $K < K_{\text{min}}$  drives the improvement. Greater improvements are achieved for higher galaxy number density or if  $k_{\text{max}}$  can be increased by a factor of 2, with milder changes if the fiducial  $b_2$  value is set to zero or shot noise on  $P_{\text{rr}}$  and  $P_{\text{gr}}$  is neglected. . . . . 162
- 5.11 As Fig. 5.9, for low-redshift (*top panels*) and high-redshift (*bottom panels*) bins of a MegaMapper-like survey. The former has greater signal to noise on reconstructed modes than DESI, leading to a greater improvement in  $\sigma(f_{\text{NL}})$  when these modes are included in the forecast. For the latter, the shot noise contributions to  $P_{\text{rr}}$  and  $P_{\text{gr}}$  are comparatively much larger, leading to a different scale-dependence for the improvement in  $\sigma(f_{\text{NL}})$ . . . . . 164
- 5.12 Modifications to the base MegaMapper forecasts. For the low-redshift bin (*left panel*), as for DESI, the improvement in  $\sigma(f_{\text{NL}})$  is driven mostly by modes of  $\delta_r$  with  $K < K_{\text{min}}$ , with further improvement possible for higher galaxy number density. For the high-redshift bin (*right panel*), neglecting  $\delta_r$  at  $K < K_{\text{min}}$  makes no difference, indicating that for lower  $K_{\text{min}}$  values, cosmic variance cancellation between  $\delta_g$  and  $\delta_r$  at the same  $K$  is driving the improvement in  $\sigma(f_{\text{NL}})$ . There are several ways to obtain greater improvements, as discussed in the main text. . . . . 164
- 5.13 As Fig. 5.9, for low-redshift (*top panels*) and high-redshift (*bottom panels*) bins of a PUMA-like survey, treating the 21 cm brightness temperature in the same way as  $\delta_g$  in our other forecasts, and translating thermal noise on the brightness temperature into an effective tracer number density for computing shot noise. We show  $\sigma(f_{\text{NL}})$  either neglecting or incorporating the effects of the 21 cm foreground wedge; at high  $z$ , the benefit to  $\sigma(f_{\text{NL}})$  from including reconstructed modes is greater in the presence of the wedge, since there are fewer  $\delta_g$  modes that can be directly measured in that case. The results for the low-redshift bin are similar to those for MegaMapper, while larger improvements in  $\sigma(f_{\text{NL}})$  are possible at higher redshift. . . . . 165

- 5.14 Modifications to the base PUMA forecasts, neglecting the foreground wedge. In the low- $z$  bin, modes of  $\delta_r$  with  $K < K_{\min}$  are entirely responsible for the improvement in  $\sigma(f_{\text{NL}})$ , with most other modifications having little effect. In the high- $z$  bin, the blue dotted curve demonstrates that the  $\sigma(f_{\text{NL}})$  improvement comes from a combination of low- $K$  modes of  $\delta_r$  and cosmic variance cancellation at higher  $K$ . The improvement would get better if the thermal noise could be reduced (which maps onto a higher  $\bar{n}$  in these forecasts). . . . . 166
- 5.15 The ratio of  $\sigma(f_{\text{NL}})$  for the  $\delta_g + \delta_r$  and  $\delta_g$ -only forecasts for the high- $z$  PUMA bin, where  $\bar{n}$  is taken to infinity and  $k_{\max}$  is artificially increased, assuming that our quadratic bias model is valid to arbitrarily high  $k$ . The left and right panels correspond to two representative values of  $K_{\min}$ . We show forecasts after marginalizing over bias parameters (*black points*) and without any marginalization (*red points*), along with the expected  $k_{\max}^{-3/2}$  scaling (*dashed lines*, each normalized to the highest- $k_{\max}$  point plotted). We find the unmarginalized curves quickly approach the ideal scaling, while the marginalized forecasts show small deviations from this scaling. This shows that large increases in constraining power are possible in principle for surveys with very high number density and a large allowed value of  $k_{\max}$ . . . . . 168
- A.1 The noise power per mode for the temperature-only estimator for different cases in the D56 region. This plot shows how the symmetric cleaned estimator presented in this work lowers the noise compared to the asymmetric estimator. The green curve shows the cross-noise between the two different asymmetric estimators, with negative values in dashed. The anti-correlation of the noise on large scales between the two different asymmetric estimators leads to a cancellation in the optimal co-add of these that results in the red noise curve for our new symmetric cleaned estimator, which recovers the forecast performance in MH18. . . . . 179
- A.2 The noise power per mode in our maps for different patches for the minimum variance (temperature + polarization) co-add of the CMB lensing maps from this work. Solid colored lines represent D56, dashed lines represent BN and dashed and dotted represent *Planck* 2018. The theory expectation for the signal is shown in black. Our maps are signal dominated for  $L < 100$  in BN and  $L < 200$  in D56. . . . . 180

B.1	Polarization weights change when accounting for the presence of biases. The solid line represents the minimum variance weights for each polarization estimator when including the presence of biases in the optimization, with $f_b = 4$ . The dashed line is the minimum variance weight when we do not consider the presence of biases. We can see that the combined $TT$ estimator becomes less important when including biases, and the $EB$ one is favored. . . . .	190
B.2	Per mode plot for the case of $f_b = 1$ for the cross power $TT$ only data optimization. On the top plot we have the signal over noise squared per mode from $TT$ , then a filtered version, with a Gaussian filter, of the absolute value of the total foreground bias for each estimator. On the bottom, the weights per mode for each estimator. . . . .	192
B.3	Per mode plot for the case of $f_b = 4$ for the cross power $TT$ only data with an LSST like sample optimization. On the top plot we have the signal over noise squared per mode from $TT$ , then a filtered version, with a Gaussian filter, of the absolute value of the total foreground bias for each estimator. On the bottom, the weights per mode for each estimator. . . . .	193
B.4	As for Figure 4.10, but now showing the case of $f_b = 1$ for the cross power $TT$ plus polarization data optimization. It can be seen that, for $f_b = 1$ where bias and variance are assigned equal importance, the optimization tries to compromise between bias and noise per mode. . . . .	193
B.5	As for Figure 4.10, but now showing the case of $f_b = 4$ for the cross power $TT$ plus polarization data optimization. It can be seen that, for $f_b = 4$ generally, the estimator with lowest bias gets selected in the combination. . . . .	194
B.6	On the left, optimization results for our baseline mask used in the main text. On the right, optimization results for a more aggressive mask, with respect to the baseline one. We can see how the results are sensitive to the used mask. . . . .	195
B.7	Validation of the pipeline for generating maps, lensing maps, and reconstructing lensing for QE, Shear, SymmPSH, SymmPH, PSH, Symm, evaluated on 80 mock maps, $20 \times 20 \text{ deg}^2$ each. . . . .	197
B.8	Biases and noises on the CMB lensing amplitude $A$ for the auto spectrum, with $TT$ only data at different lmaxes of reconstruction. The bands bands represent regimes where the bias is greater than a fraction of the noise. We can see in action the bias-noise trade off: as we increase the noise, the biases get reduced, and vice versa. . . . .	200

- B.9 Biases and noises on the CMB lensing amplitude  $A$  for the auto spectrum, with  $TT$  plus (foreground free) polarization data, for the estimators presented in the main text, at different  $l_{\max}$ es of reconstruction. Thanks to how clean the polarization is, we are easily able to reduce the bias without a high cost in SNR. . . . . 200
- B.10 Same as Figure B.8 for  $TT$  with a more aggressive masking. In general, the aggressive masking will have more impact on the trispectrum term of the foreground induced CMB lensing bias. This will lead to a change in hierarchy among the biases at different  $l_{\max,TT}$  for the same estimator, e.g. for PH going to  $l_{\max,TT} = 4000$  leads to a lower bias with respect to  $l_{\max,TT} = 3000$ , . . . . . 201
- B.11 Biases and noises on the CMB lensing amplitude  $A$  for the cross spectrum with an LSST-like sample, with  $TT$  only data, for the estimators presented in the main text, at different  $l_{\max}$ es of reconstruction. Here, the bias barely reaches the regime of  $b < \sigma/2$ . Therefore, for this case an optimization is needed to reach a subdominant bias with respect to the noise. . . . . 201
- B.12 Same as Figure B.11 for the cross spectrum with an LSST-like sample, for  $TT$  plus (foreground free) polarization. Thanks to how clean the polarization is, we are easily able to reach a regime where the bias is subdominant compared to the noise for several estimators at several  $l_{\max,TT}$ -es, as the cross-primary term goes as  $\alpha_T T^2$ . . . . . 202
- B.13 Total relative foreground-induced auto-spectrum CMB lensing bias per mode for each estimator considered in this work, for  $l_{\max} = 3500$ . In the top panel is shown the contribution arising when using the point source mask constructed on 148 GHz data; in the bottom panel, when using the point source mask constructed from the product of point source masks on the SO frequencies. When calculating the ILC combinations we fix the weights to theory ones, without considering the change in mask. We can see how the aggressive masking reduces the biases, with particular emphasis on the QE, meaning that the trispectrum-foreground-induced bias term is the most affected by the masking operation. . . . . 204
- B.14 Same as Figure B.13 but for the cross-spectrum of CMB lensing with an LSST-like sample. We can see how the aggressive masking reduces the biases, with particular emphasis on the QE. . . . . 205

- C.1 Cross-correlation coefficients between G, S, and T estimators, as given by Eq. (C.44), for our DESI-like survey forecast (the other surveys give similar results) We see that all three estimators are highly correlated, implying, through Eq. (C.46), that applying bias-hardening will strongly increase the noise of the resulting estimator. . . . . 218
- C.2 We plot the contamination curves from numerical and analytic approximation. It can be seen that for the  $\varphi\varphi$ ,  $b_{11}$  and  $b_{02}$  terms we have a  $\frac{1}{K^2}$  behaviour. For the analytical calculation of the  $b_{02}$  term we use an approximation  $\left( \int_{q_{\min}}^{q_{\max}} dq q^2 \frac{P_{\text{lin}}^2(\vec{q})}{P_{\text{NL}}^2(\vec{q})} \frac{1}{M(\vec{q})} / \int_{q_{\min}}^{q_{\max}} dq q^2 \frac{P_{\text{lin}}^2(\vec{q})}{P_{\text{NL}}^2(\vec{q})} \right) \approx \frac{1}{2M(q_{\min})}$ . We also show the absolute value of the numerical  $b_{01}$  curve, with an approximate constant value on large scales. . . . . 225
- C.3 The fraction of small-scale modes lying outside the foreground wedge, for the low- $z$  (*solid*) and high- $z$  (*dashed*) PUMA redshift bins we use in our forecasts. To a good approximation, the reconstruction noise integrals  $N_{\alpha\beta}$  will simply be scaled by the inverse of these fractions in the presence of the wedge. . . . . 226
- C.4 Forecasts for a PUMA-like survey, analogous to Fig. 5.13 except for using a cutoff on the line-of-sight component of accessible  $\delta_{\text{g}}$  modes instead of an isotropic  $K$  cutoff. This is motivated by the fact that 21 cm foregrounds will preferentially obscure modes with low wavenumber components along the line of sight. The absolute values of  $\sigma(f_{\text{NL}})$  are slightly higher when using  $K_{\parallel, \text{min}}$  instead of  $K_{\text{min}}$ , since more modes are eliminated with a  $K_{\parallel}$  cut, but the improvement in  $\sigma(f_{\text{NL}})$  from including reconstructed modes is qualitatively similar to the case with  $K_{\text{min}}$ . . . . . 228

## Introduction

We live in the era of big data. Current and future cosmological surveys will probe the Universe with unprecedented precision. We will challenge again the simple and, until now, very successful, mainstream cosmological model, and in doing so hope to gain insights into its unknown pillars: dark matter, dark energy, and inflation.

Gone are the days where mysteries could be unraveled with little thought. Cautious and ambitious has to be he who wants to take paths from theory to observations. Ambitious, to get the most out of new cosmological data. Cautious, as surveys are no longer statistically limited, as we have to consider more systematic effects.

The focus of this thesis is to provide methods for more robust and enhanced reconstructions of the sky, or more precisely, the large-scale structure.

The distribution of matter in the Universe weakly lenses the cosmic microwave background (CMB) photons along their way to us. In the last decade, several surveys of the sky used small-scale CMB information to reconstruct the projected matter distribution, and extract insights about dark energy, neutrino masses, and dark matter.

As we go to smaller and smaller CMB scales, foregrounds dominate more and more. This might result in biased measurements that lead to an incorrect cosmological inference. For low-resolution surveys, this was not significant; however, for high-resolution temperature-dominated ones, foreground contamination might induce detectable biases. The first two parts of this thesis focus on exploring and implementing methods for foreground reduction. We quantify foreground biases arising in CMB lensing analyses. Then, we develop methods for foreground mitigation, preserving most of the signal-to-noise.

Once we have a clean, robust CMB lensing map, we can cross-correlate it with large-scale structure probes. Cross-correlation science has become one key way to constrain cosmology. Galaxies and CMB lensing trace the same common underlying matter field. Cross-correlations break degeneracies among certain parameters, cancel

---

systematics in auto-analyses, allow for sample variance cancellation, and make possible tomographic analyses to trace the growth of structure with time. Using the methods developed in this thesis, we will build a foreground-free CMB lensing map, which we will then cross-correlate with BOSS galaxies.

In principle, we should be able to cross-correlate CMB lensing with any other probe of the matter, but in practice this is not always possible. For example, intensity mapping has a good overlap in the high-redshift range with CMB lensing, but the foreground filtering of intensity mapping surveys removes most of the low- $k_{\parallel}$  modes that source CMB lensing, destroying the correlation between the two observed fields. A possible solution is to get the lost intensity mapping large-scale modes from small-scale power.

Due to gravity, long-wavelength density modes affect the power of the locally measured small-scale matter power. Exploiting this, it is possible to reconstruct large-scale modes from small-scale ones. Inspired by CMB lensing reconstruction, in the last chapter we present a method to reconstruct large-scale modes from galaxy surveys, obtaining a new probe for cosmology. Our method accounts for the effects of non-linear bias and non-Gaussianity for the first time. We perform a Fisher forecast for a joint constraint primordial local non-Gaussianity with this new field and the original one. By taking advantage of cosmic variance cancellation, we show that we can improve our constraints on  $f_{\text{NL}}$ . This will be especially relevant for upcoming large-scale structure observations. It might help reduce viable inflationary theories and guide us into our journey to obtaining a deeper understanding of the physical world.

## Overview of this thesis

Formally the thesis is structured as follows.

1. In Chapter 2 we start by reviewing some background cosmology. We discuss primordial inflationary perturbations, subsequent generation of CMB perturbations, and some secondary anisotropies of the CMB. We devote time to CMB lensing, followed by a presentation of dark matter perturbations. We end discussing the topic of large-scale reconstruction with quadratic estimators.
2. In Chapter 3 we present a cross-correlation between ACTPol CMB lensing maps and CMASS BOSS galaxies ([72]). The CMB lensing maps are built with a novel foreground-free estimator which exploits multi-frequency cleaned data.
3. In Chapter 4 we first give an overview of some known CMB lensing estimators for foreground mitigation. Then, we optimally combine them, with respect to some

loss function, to extract the best possible combination that minimizes foreground induced bias and maximizes the signal-to-noise, for the lensing amplitude in auto and cross-correlations.

4. In Chapter 5, we discuss a quadratic estimator-based reconstruction of the linear density field from galaxy clustering. We incorporate non-linearities due to gravity, galaxy biasing, and primordial non-Gaussianity, then verify our predictions with  $N$ -body simulations. We perform a Fisher matrix analysis to show that the reconstructed field in combination with the biased tracer field can improve constraints on local-type primordial non-Gaussianity. We find significant improvement on constraints due to cosmic variance cancellation, resulting from the additional correlated modes of the reconstructed field, similar to multi-tracer analyses ([71]).
5. Finally, in Chapter 6 we present a summary of our results and the impact these results have in the field, including a future outlook.

## Review of Cosmological Perturbations

Our goal in this chapter is to understand cosmological perturbations. We start with some background cosmology. We then proceed to a discussion of how to describe cosmological observables. Inflationary cosmology, as well as CMB anisotropies are considered. The focus then shifts to CMB secondary anisotropies, with particular emphasis on CMB lensing. We also introduce perturbation theory for large-scale structure and biasing. Finally, we discuss large-scale reconstruction from CMB and large-scale structure.

### 2.1 The Expanding Universe

We will use units such that  $c = 1$ . The metric signature is  $(-, +, +, +)$ .

#### 2.1.1 The FLRW Metric

We assume, on large scales, a homogeneous and isotropic space evolving in time. We write the most general metric satisfying these geometrical considerations in this way (e.g. [50])

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu = -dt^2 + a^2(t) \gamma_{ij} dx^i dx^j, \quad (2.1)$$

where  $g_{\mu\nu}$  is the four metric,  $\gamma_{ij}$  represents the metric of the space,  $t$  is the proper time,  $x^i$  are comoving coordinates, and  $a$  is the scale factor. We slice the space time into maximally symmetric (i.e. homogeneous and isotropic) spatial slices  $\Sigma$  rescaled by the scale factor as a function of time. Latin indices  $(i, j, \dots)$  represent spatial components, and Greek indices  $(\mu, \nu, \dots)$  all four space time components. The inverse of the four-metric  $g^{\mu\nu}$  raises space time indices (and likewise  $\gamma^{ij}$ , the inverse of  $\gamma_{ij}$ , for space indices).

The metric in (2.1) is also known as FLRW metric, and it can be written as

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu = -dt^2 + a^2(t) \left[ \frac{dr^2}{1 - kr^2} + r^2 d\Omega^2 \right], \quad (2.2)$$

where  $k$  is a curvature parameter,  $r$  a distance coordinate with units of length, and  $d\Omega^2 = d\theta^2 + \sin^2\theta d\phi^2$  is the metric on the two-sphere. If  $k = 0$  we have flat space,  $k > 0$  a three-sphere, and  $k < 0$  a hyperbolic space. In this thesis we focus on flat space, which has Cartesian coordinates  $\gamma_{ij} = \delta_{ij}$ . In this case the distance coordinate  $r$  represents a radius.

We can use the metric discussed for several purposes, one of which is studying the propagation of light. Light rays propagate on null geodesics,  $ds^2 = 0$ .<sup>1</sup> Light trajectories are simple in certain coordinates. Let us switch our time variable from physical time to conformal time:

$$d\eta \equiv \frac{dt}{a(t)} \quad (2.3)$$

such that the FLRW line element becomes

$$ds^2 = a^2(\eta) \left[ -d\eta^2 + \gamma_{ij} dx^i dx^j \right]. \quad (2.4)$$

Due to isotropy, the radial trajectory with  $\phi, \theta = \text{const.}$  is a geodesic, and  $d\eta^2 = \frac{dr^2}{1 - kr^2} \equiv d\chi^2$ , and radial geodesics are described by  $\chi(\eta) = \pm\eta + \text{const.}$ , corresponding to simple straight lines at  $\pm\frac{\pi}{4}$  in the  $\eta$ - $\chi$  plane. These will set the limit of how far we can see.

If the universe has a finite age, starting at some time  $t_i$ , then there is a maximum amount of time for light to have traveled. The volume of space from which we can receive information at any given time is limited. The boundary of the volume is called the particle horizon [191]. The maximum comoving distance light can travel is,

$$\chi_p(\eta) = \int_{t_i}^t \frac{dt}{a(t)} = \eta - \eta_i, \quad (2.5)$$

where  $\eta_i$  sets the beginning of the universe.

To obtain the physical distance between two radially separated points on the same spatial slice, parametrized by  $\eta$ , we multiply  $\chi_p$  with the scale factor:

$$d_p(t) = a(\eta) \chi_p(\eta). \quad (2.6)$$

---

<sup>1</sup>This is true in special relativity. Then, it is true in general relativity for local inertial coordinate frames. Given the invariance of the line element under coordinate transformations, this is true for light geodesics in curved space times.

In practice, the maximum distance light can travel to us in our universe is set by the time of recombination. Suppose we shoot out a photon today from our position, making it “go back” in time. It will scatter frequently around 13.7 billion years ago, because prior to that, the Universe was opaque to photons. It trapped them in a plasma, and released them near hydrogen recombination. Thus, there is a maximum distance that we can see in practice with light. This is the “optical” horizon  $d_{\text{opt}} = a(\eta)(\eta - \eta_r) = a(t) \int_{t_r}^t \frac{dt}{a}$ , where  $\eta_r$  ( $t_r$ ) is the time of decoupling [191]. In practice we can infer information about earlier times with these photons. In principle we could directly look back with cosmic background neutrinos or cosmological gravitational waves, as they decouple earlier. In practice this is difficult for now, although CMB perturbations still allow us to infer information about the very early universe.

To characterize distances and time traditionally we use the redshift. Redshift is the fractional shift in the wavelength of a photon emitted from someplace at time  $t_e$  and observed today at  $t_0$ :

$$z = \frac{\lambda_0 - \lambda_e}{\lambda_e}, \quad (2.7)$$

where  $\lambda_0$  is the wavelength today, and  $\lambda_e$  is the wavelength at emission.

The expansion of the Universe leads to a cosmological redshift of the photon wavelength,  $\lambda \propto a(\eta)$ , so that

$$1 + z = \frac{a_0}{a_e}. \quad (2.8)$$

We can see that there is a one-to-one relation between  $z$  and  $a_e \equiv a(t_e)$ , making the redshift a measure of time, if we fix  $t_0$  for  $a_0 \equiv a(t_0)$ . On the other hand, the redshift can also be used as a measure of space. Indeed, the comoving distance

$$\chi = \eta_0 - \eta_e = \int_{t_e}^{t_0} \frac{dt}{a(t)} = \frac{1}{a_0} \int_0^z \frac{dz'}{H(z')} \quad (2.9)$$

can be written in terms of the redshift, where we used that  $dz = -(1+z)H(t)dt$  by differentiating the definition of redshift,  $\chi = r$  for a flat universe, and  $H(t) = \frac{1}{a} \frac{da}{dt} \equiv \frac{\dot{a}}{a}$  the Hubble parameter as a function of time.

We are often interested in measuring the angular size of observed objects. The angular size is the angle that an object of given proper physical size subtends on the sky. The same object can have different angular sizes, depending on the underlying cosmology. Relating a proper physical size to the observed one, we can extract cosmological parameters.

One definition of distance in cosmology comes from the following. In a Euclidean space, an object of proper size  $s$ , subtending a small angle  $\theta$ , has  $s = a(t_e)\chi\theta$ , where

$\chi$  is the comoving radial distance from the observer. For general geometries, this is not true, as the geometry of the universe might not be flat. We define the angular diameter distance  $d_A$  to be the distance such that  $\theta$  can still be given by the Euclidean relation [284]

$$\theta = s/d_A , \quad (2.10)$$

where in flat space

$$d_A = a(t)\chi . \quad (2.11)$$

The evolution of  $a(t)$  is important for understanding the FLRW metric.

To describe the evolution of the geometrical properties of space time described by the metric, and in FLRW by  $a$ , we need to solve Einstein's equations,

$$G_{\mu\nu} = 8\pi GT_{\mu\nu} + \Lambda g_{\mu\nu} , \quad (2.12)$$

where  $G_{\mu\nu}$  is the Einstein tensor, related to the space time Ricci tensor  $R_{\mu\nu}$  (depending on the metric and its derivatives):

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R , \quad R \equiv g^{\mu\nu}R_{\mu\nu} , \quad R_{\mu\nu} \equiv R^{\lambda}_{\mu\lambda\nu} , \quad (2.13)$$

with the space time Riemann tensor defined as

$$R^{\mu}_{\nu\kappa\lambda} = \partial_{\kappa}\Gamma^{\mu}_{\nu\lambda} - \partial_{\lambda}\Gamma^{\mu}_{\nu\kappa} + \Gamma^{\mu}_{\alpha\kappa}\Gamma^{\alpha}_{\nu\lambda} - \Gamma^{\mu}_{\alpha\lambda}\Gamma^{\alpha}_{\nu\kappa} , \quad (2.14)$$

and the affine connection coefficients  $\Gamma^{\mu}_{\nu\kappa}$

$$\Gamma^{\mu}_{\nu\kappa} = \frac{1}{2}g^{\mu\lambda}(\partial_{\nu}g_{\kappa\lambda} + \partial_{\kappa}g_{\lambda\nu} - \partial_{\lambda}g_{\nu\kappa}) . \quad (2.15)$$

The the energy-momentum tensor  $T_{\mu\nu}$  describes the content of the Universe. We will choose to model it with a perfect fluid description:

$$T_{\mu\nu} = (p + \rho)u_{\mu}u_{\nu} + pg_{\mu\nu} , \quad (2.16)$$

where  $p$  and  $\rho$  are the proper pressure and energy density measured in the fluid rest frame, and  $u^{\mu} = \frac{dx^{\mu}}{d\lambda}$  ( $d\lambda^2 \equiv -ds^2$ ) is the fluid four velocity (e.g. [39]). Comoving coordinates will be chosen such that the fluid is at rest,  $u^i = 0$ ,  $\forall i \in \{1, 2, 3\}$ , and we see that  $T^{00} = \rho$ ,  $T^{0i} = 0$ , and  $T^{ij} = a^{-2}p\delta^{ij}$ .<sup>2</sup>

<sup>2</sup>If to describe the components in the Universe there are other effects like weak collisions, or thermal conduction, then  $T^{\mu\nu}$  acquires an extra term  $\Sigma^{\mu\nu}$  to describe the imperfectness of the fluid (e.g. [39, 284]).

### The Friedmann Equation

For the FLRW metric, Einstein's equations, with a perfect fluid, reduce to two equations (by symmetries of space time). The Friedmann equation comes from the 00 component:

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2} + \frac{\Lambda}{3} ; \quad (2.17)$$

and the second Friedmann, or Raychaudhuri, equation comes from the  $ii$  ones:

$$\frac{\ddot{a}}{a} = -\frac{4\pi}{3}G(\rho + 3p) + \frac{\Lambda}{3} , \quad (2.18)$$

where  $H = \frac{\dot{a}}{a}$  is the Hubble parameter. The energy density appearing in both equations can be understood as the sum of (independent) fluid components  $i$ , with their own energy density,  $\rho_i$  and pressure,  $p_i$ . Therefore,  $\rho$  is to be understood as  $\rho = \sum_i \rho_i$ . The dominant components will vary in time. For example, in the early Universe, assuming negligible contribution from dark energy in the form of cosmological constant, we have  $\rho = \rho_b + \rho_\gamma + \rho_m + \rho_\nu$ , where we have baryon, photon, dark matter, and neutrino contributions respectively. For now, we will put this aside, and write just  $\rho$ .

A relativistic perfect fluid will obey a local continuity equation (in the absence of collision terms)

$$\nabla_\mu T^{\mu\nu} = 0 . \quad (2.19)$$

From the continuity equation for FLRW with a perfect fluid

$$\dot{\rho} = -3H(\rho + p) , \quad (2.20)$$

which is just a manifestation of Bianchi's identity, as it can be derived from the Friedman and Raychadhuri equations.

To solve for the evolution of  $a, \rho, p$ , we also need to specify an equation of state connecting the density  $\rho$  to the pressure  $p$ . We will take the simple relation of a barotropic fluid  $p = p(\rho) = w\rho$ , where  $w$  will be assumed constant.<sup>3</sup> For example, the case with  $w = 0$  represents pressureless dust, a good approximation for collisionless non-relativistic fluids.

The continuity equation then becomes

---

<sup>3</sup>This is not a valid assumption for all components. For example, from a phenomenological point of view, it is possible to model the late-time dark energy with  $w = w_0 + (1 - a)w_a$ , with  $w_0, w_a$  some parameters, and in this case, the equation of state becomes time-dependent.

$$\frac{\dot{\rho}}{\rho} = -3H(1+w) . \quad (2.21)$$

Solving for  $\rho$  we find  $\rho \propto a^{-3(1+w)}$ , and, for  $w \neq -1$ ,  $a \propto t^{\frac{2}{3(1+w)}}$  or  $a \propto \eta^{\frac{2}{1+3w}}$ .

The specific behavior of  $a$  then depends on the specific components that dominate the Universe. The most important ones are: photons ( $\gamma$ ) and (relativistic/massless) neutrinos ( $\nu$ ) as radiation with  $w = \frac{1}{3}$ , electrons ( $e$ ) and protons ( $p$ ) as baryons ( $b$ ), cold dark matter ( $c$ ) that with baryons constitute matter ( $m$ ) with  $w \approx 0$ , and finally dark energy ( $\Lambda$ ), with  $w = -1$  for a cosmological constant.<sup>4</sup>

In cosmology, it is useful to define the critical density  $\rho_c \equiv \frac{3H^2}{8\pi G}$ . Then, for any component  $\rho_i$  we define the density parameter  $\Omega_i(a) = \rho_i(a)/\rho_c(a)$ , and  $\Omega_k(a) \equiv -\frac{k}{H^2(a)a^2}$ . When the argument of  $\Omega_i(a)$  is not written,  $\Omega_i$  will mean its present value,  $\Omega_i \equiv \Omega_i(a_0)$ , where  $a_0$  is the scale factor today, chosen such that  $a_0 = 1$ .<sup>5</sup>

Then, we can write the Friedmann equation in the useful form

$$1 = \sum_{i \in \mathcal{C}} \Omega_i(a) + \Omega_k(a) \quad (2.22)$$

In particular, for a flat universe, the sum of the densities over the components  $\mathcal{C}$  satisfies  $\sum_{i \in \mathcal{C}} \Omega_i(a) = 1$ , and we see that the critical energy density is quite useful, as in this case, it is equal to the total energy density matching the Hubble constant.

Today, we know that the energy density budget of the Universe is given by roughly 31% matter, of which 5% baryonic matter and the rest is dark matter, 69% dark energy. Other contributions are from photons  $\Omega_\gamma \approx 5 \times 10^{-5}$ , and neutrinos  $\Omega_\nu \sim \mathcal{O}(10^{-3})$ .<sup>6</sup> Finally, we have  $|\Omega_k| < 0.01$ .

To derive these numbers, one has to go through a cosmological analysis, where observables are studied to connect to the fundamental parameters. Thus, we proceed now to discussing how we describe cosmological observables.

<sup>4</sup> $\Lambda$  is used for the cosmological constant, but as our current cosmological model is based on that, we will stray from this, and refer to dark energy in general with  $\Lambda$ .

<sup>5</sup>In equation (2.2) we can rescale  $a \rightarrow \lambda^{-1}a$ ,  $r \rightarrow \lambda r$ ,  $k \rightarrow \lambda^{-2}k$ , without changing it. Therefore we can always choose  $\lambda$  to be the value of  $a$  today, such that with the rescaling  $a_0 = 1$ .

<sup>6</sup>Neutrinos are considered radiation in the early universe, where they are relativistic. As the universe expands and they slow, they become non-relativistic and therefore count as matter. The value of  $\Omega_\nu \sim \mathcal{O}(10^{-3})$  comes from considering the neutrinos non-relativistic with some mass: roughly  $n_\nu m_\nu$ , where  $n_\nu$  is the number density of neutrinos, and the  $m_\nu$  the sum of their masses.

## 2.2 How do we describe cosmological observables?

We assume that the density variations that we observe today in the Universe are from a random process whose statistics, but not any of its particular realizations, can be predicted. This can be described by a random field in space  $f(\vec{x})$  with a functional  $\mathcal{P}(f)$ , indicating the probability of realizing some field configuration of  $f$  in space.

A basic quantity is the correlator, defined as

$$\langle f(\vec{x})f(\vec{y})\dots \rangle = \int \mathcal{D}f \mathcal{P}(f) f(\vec{x})f(\vec{y})\dots, \quad (2.23)$$

where  $\langle \dots \rangle$  denotes an ensemble average, an average over many realizations of the random process. This definition can be generalized to include multiple different fields, although we will stick to only one here. Correlators are useful as they can be used to study the probability functional describing  $f$ .

In accordance with observations, we will assume  $f$  to have two properties:

- Statistical isotropy, in the sense that the correlation (2.23) is invariant under rotations. The rotated field  $\hat{R}f(\vec{x}) = f(R^{-1}\vec{x})$  has the same expectation value (2.23) as  $f$  for any rotation  $R$  in space described by the group representation  $\hat{R}$  on the space of functions.
- Statistical homogeneity, in the sense that the correlation (2.23) is invariant under translations. The translated field  $\hat{T}f(\vec{x}) = f(\vec{x} + \vec{a})$  has the same expectation value (2.23) as  $f$ , for any shift  $\vec{a}$  described by the operator  $\hat{T}$ .

Under homogeneity, and assuming that in a correlator the  $f$ s with distant arguments are uncorrelated (i.e.  $\lim_{|\vec{u}| \rightarrow \infty} \langle f(\vec{x} + \vec{u})f(\vec{y} - \vec{u})\dots \rangle = \lim_{|\vec{u}| \rightarrow \infty} \langle f(\vec{x})\dots \rangle \langle f(\vec{y})\dots \rangle$ ), it is possible to prove the ergodic theorem [284], which implies that

$$\langle f(\vec{x})f(\vec{y})\dots \rangle = \int \mathcal{D}f \mathcal{P}(f) f(\vec{x})f(\vec{y})\dots \approx \frac{1}{V} \int_V d^3\vec{z} f(\vec{x} + \vec{z})f(\vec{y} + \vec{z})\dots \quad (2.24)$$

where  $V$  is a large enough volume such that spatial correlations can decay sufficiently rapidly with increasing separation in such a way that we have many statistically independent volumes in one realization [284]. Each volume should provide a fair sample of the original (functional) probability distribution describing  $f$ .

To recap, ergodicity means that the ensemble average (2.23) is equal to a spatial average taken over one realization of the random field over a large volume in our universe realization. This is important, as it will allow us to make (volume) averages from observations, and connect these observations to theory.

If  $V$  is not large enough, the spatial average can be quite different from the ensemble expectation value. The mean square difference between the ensemble expectation value of a quantity predicted by theory and the observed average approximating it is called cosmic variance [167, 284]. This limits how accurately we can compare theory and observations. We will see examples later for the CMB, as well as a way we can overcome these limits in Chapter 5.

Given a volume, and a large number of independent, identically distributed random fields, by the central limit theorem, the distribution of their sum tends to a Gaussian.<sup>7</sup>

### Gaussian distributions

Let us consider  $f$  with zero ensemble mean.<sup>8</sup>

The functional governing  $f$  is said to be Gaussian if the average of a product of an even number of  $f$ s is given by the sum over averages of pairs of all the different ways of pairing the  $f$ s with each other [284]

$$\langle f(\vec{x})f(\vec{y})\dots \rangle = \sum_{\text{ordered pairings}} \prod_{\text{unique pairs}} \langle ff \rangle, \quad (2.25)$$

and any expectation value of an odd number of  $f$ s vanishes. Put simply, the two-point function, and the mean, describes the entire field.<sup>9</sup> As an example, the CMB can be assumed to not deviate from a Gaussian [11]. As the Ergodic theorem applies to Gaussian random fields (assuming  $\lim_{|\vec{u}| \rightarrow \infty} \langle f(\vec{x} + \vec{u})f(\vec{y} - \vec{u})\dots \rangle = \lim_{|\vec{u}| \rightarrow \infty} \langle f(\vec{x})\dots \rangle \langle f(\vec{y})\dots \rangle$  sufficiently quickly) [284], we then simply need to take averages over data to study it.

### Non-Gaussian distributions

If we find any extra residual in the correlation functions with respect to what defines a Gaussian field, then we have a non-Gaussian field. There are many ways to be non-Gaussian. A simple one is by squaring a Gaussian field  $f$ , as  $g \equiv f^2$  will have a non zero  $\langle ggg \rangle$ . There are scenarios where non-Gaussianity could arise in the early universe [167], although from current observations, primordial fluctuations are still constrained to be nearly Gaussian (e.g. [11]).

Later, we will see that the additional presence of correlation in the four-point correlator of the CMB, with respect to a Gaussian-only case, leads to additional powerful observables in cosmology.

<sup>7</sup>In general it is important that the sum is not dominated by one variable, or a small, number of variables. If we have equally distributed variables, then this is satisfied [85].

<sup>8</sup>If there is any non-zero ensemble mean we just subtract it from  $f$ .

<sup>9</sup>The definition that we provided for a Gaussian random field is basically the Isserliss-Wick theorem (e.g. [85]).

### 2.2.1 Fourier space

Cosmological fields can be described in real space, and in other spaces, such as Fourier space. Each one has its own advantages and disadvantages. In particular, the Fourier space description is useful as it focuses on the behavior on a certain scale (e.g. [95]).

Given a field  $f$ , we consider the Fourier transform, with the following convention

$$f(\vec{k}) \equiv \tilde{f}(\vec{k}) = \int d^3\vec{x} \exp[-i\vec{k} \cdot \vec{x}] f(\vec{x}) \quad (2.26)$$

with an inverse

$$f(\vec{x}) = \int \frac{d^3\vec{k}}{(2\pi)^3} \exp[i\vec{k} \cdot \vec{x}] f(\vec{k}) . \quad (2.27)$$

We can make a translation by  $\vec{a}$ , so that the translation representation in the Fourier space gives an extra phase factor for the Fourier transform

$$f(\vec{k}) \rightarrow \exp[i\vec{k} \cdot \vec{a}] f(\vec{k}) , \quad (2.28)$$

and similarly for the product of two fields

$$f(\vec{k}) f^*(\vec{k}') \rightarrow \exp[i(\vec{k} - \vec{k}') \cdot \vec{a}] f(\vec{k}) f^*(\vec{k}') . \quad (2.29)$$

Taking the ensemble average of this, and requiring invariance under translations, we see that the exponential term above requires  $\vec{k}' = \vec{k}$ . Therefore we require [54, 21]

$$\langle f(\vec{k}) f^*(\vec{k}') \rangle = (2\pi)^3 \delta_D^{(3)}(\vec{k} - \vec{k}') F(\vec{k}) . \quad (2.30)$$

Similarly, by requiring statistical invariance under rotations we obtain that [21]

$$\langle f(\vec{k}) f^*(\vec{k}') \rangle = (2\pi)^3 \delta_D^{(3)}(\vec{k} - \vec{k}') F(|\vec{k}|) . \quad (2.31)$$

If a field is real,  $f(\vec{k}) = f^*(-\vec{k})$ , the expression above can also be written as  $\langle f(\vec{k}) f(\vec{k}') \rangle = (2\pi)^3 \delta_D^{(3)}(\vec{k} + \vec{k}') F(|\vec{k}|)$ .

Working in Fourier space allows for certain operations, such as smoothing with some window function  $W$ :

$$f_R(\vec{x}) = \int d^3\vec{y} W_R(\vec{y}) f(\vec{x} - \vec{y}) \rightarrow f_R(\vec{k}) = f(\vec{k}) W_R(\vec{k}) , \quad (2.32)$$

where we have taken advantage of the convolution theorem

$$f(\vec{x}) = [g * h](\vec{x}) = \int d^3\vec{y} g(\vec{y}) h(\vec{x} - \vec{y}) \rightarrow f(\vec{k}) = g(\vec{k}) h(\vec{k}) . \quad (2.33)$$

which is important for many numerical applications.

In general  $f$  will represent an observable,  $\mathcal{O}$ , with some perturbation with respect to a background value,  $\bar{\mathcal{O}}$ . In cosmology, perturbations are a tricky concept which require attention here.

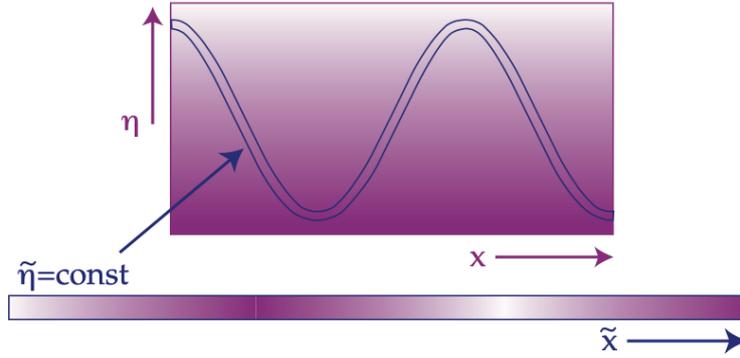
### 2.2.2 Cosmological perturbations: gauge issue

Thus far, we have focused on a homogeneous universe, but our universe contains large-scale structure, such as galaxies and galaxy clusters, as a consequence of the growth of small-in-amplitude density fluctuations in the early universe.

We used the term small-in-amplitude density fluctuations. With this we imply that the early universe space time is close to the homogeneous and isotropic one.

Suppose we have some quantity  $Q$  in the physical space time. The perturbation of this quantity is defined as its value in the physical space time with respect to its value in the hypothetical unperturbed reference background, at the same coordinate values. Since we have two different space times, we need to specify some mapping between them in order to obtain this difference, also known as gauge choice. In general relativity there is no unique choice for a mapping, as we are free to choose any coordinate system. This introduces gauge artifacts (e.g. [191, 284]), where even a homogeneous space time can have fluctuations with a coordinate change.

As an example, suppose we have a homogeneous isotropic fluid described by  $\rho(\eta, \vec{x}) = \bar{\rho}(\eta)$ . General relativity allows us to choose any coordinate system by specifying hypersurface slices with constant-time, and line threads with fixed-space. Let us take a position-dependent time slicing, where  $\eta \rightarrow \tilde{\eta} = \eta + \xi(\eta, \vec{x})$ . Then,  $\bar{\rho}(\eta) \rightarrow \tilde{\rho}(\tilde{\eta}, \vec{x}) \equiv \bar{\rho}(\eta(\tilde{\eta}, \vec{x})) = \bar{\rho}(\tilde{\eta} - \xi) = \bar{\rho}(\tilde{\eta}) - \bar{\rho}'(\tilde{\eta})\xi(\eta, \vec{x}) \equiv \bar{\rho}(\tilde{\eta}) + \delta\bar{\rho}(\tilde{\eta}, \vec{x})$ , for a small  $\xi$ . We can see that our total density is given by a background value plus a position dependent perturbation, even if that perturbation itself is not physical in origin. Figure 2.1 illustrates visually the implication of this.



*Fig. 2.1* A one-dimensional density field evolving in time. On the  $y$ -axis, is the time slice, on the  $x$ -axis, the position. We can see that the field is spatially constant in these coordinates. If we now choose another time slicing, represented by  $\tilde{\eta}$ , we see that for the new time coordinate  $\tilde{\eta}$ , the field is no longer constant. This figure is taken from [121].

To resolve this discrepancy between real and fake perturbations, one could identify gauge-invariant quantities [27, 190], or fix a gauge and perform calculations in this. We will see an example of this when studying the CMB fluctuations and inflation.<sup>10</sup>

## 2.3 How perturbations may arise

Cosmological inflation is one of the most important theories of the early Universe. It addresses problems of the old big bang model, such as why the Universe is so homogeneous on large scales. Furthermore, it has a mechanism for seeding the structure we observe today in the Universe.

There are three important generic predictions for the properties of perturbations generated by (slow-roll single field) inflation: Gaussianity, adiabaticity, and nearly scale-invariant initial conditions. Until now, we have still not seen deviations from these [13].

<sup>10</sup>There are advantages and disadvantages to working with different gauges. For example, for the scalar perturbations in inflation, working in the Newtonian gauge, a good choice for CMB perturbations, mixes the spatial metric perturbation  $\Phi$  and the inflaton perturbation  $\delta\phi$ , making equations difficult. Switching to the so-called spatially flat slicing makes calculations much easier, so one may prefer to work in this one. Gauge-invariant variables can help connecting between the two gauges.

### 2.3.1 Seeding perturbations with inflation

Inflation solves several problems of the hot Big Bang paradigm. The problems are as follows (e.g. [284]): large-scale homogeneity in the structure observed today, despite the absence of causal contact between very distant regions (horizon problem); the flatness of the Universe, despite the instability of the spatial curvature term during matter/radiation domination (flatness problem); the absence of magnetic monopoles left over from phase transitions at very high energies (monopole problem).

Inflation is a period of rapid accelerated expansion in the early Universe which lasts long enough to solve these problems (e.g. [284, 81]).

This process can be described by a slowly-varying vacuum energy component such that the scale factor  $a(t)$  grows quasi-exponentially. This means that a magnetic monopole ( $\rho \propto a^{-3}$ ) will be washed out, losing energy density faster than the source of inflation, and the same happens for the spatial curvature term (with a corresponding equivalent energy density  $\rho \propto a^{-2}$ ). The beauty of inflation is that it also provides a mechanism for the generation of the primordial fluctuations that seeded the structure we observe today in the Universe, such as galaxy clusters or CMB fluctuations.

The important aspects of inflation can be shown with a single scalar field  $\phi(\vec{x}, t)$ , the inflaton field, with a large but flat potential  $V(\phi)$ , with an action given by

$$S = \int d^4x \sqrt{-\text{Det}(g)} \left[ \frac{M_{\text{pl}}^2 R}{2} - \frac{1}{2} g^{\mu\nu} \nabla_\mu \phi \nabla_\nu \phi - V(\phi) \right]. \quad (2.34)$$

By varying the action, the equation of motion of this field in the FLRW metric is given by

$$\ddot{\phi} + 3H\dot{\phi} - \frac{\nabla^2 \phi}{a^2} = -\frac{\partial V}{\partial \phi}. \quad (2.35)$$

If the scalar field is homogeneous and isotropic, then the stress-energy tensor is described by [30]<sup>11</sup>

$$\rho_\phi = \left( \frac{1}{2} \dot{\phi}^2 + V(\phi) \right), \quad p_\phi = \left( \frac{1}{2} \dot{\phi}^2 - V(\phi) \right). \quad (2.36)$$

To have inflation we need an accelerated expansion  $\ddot{a} > 0$ , and from the Raychaudhuri equation, we see that this requires  $\rho + 3p < 0$ . If the potential is large enough compared to the kinetic term, i.e.  $V(\phi) \gg \frac{1}{2} \dot{\phi}^2$ , then  $p_\phi \approx -\rho_\phi$ . For a flat large enough potential, we will therefore have cosmological constant-like behavior. Importantly in this case  $a \propto e^{Ht}$ , and the ordinary matter will scale as  $\rho_m \propto a^{-3}$ , while radiation  $\rho_\gamma \propto a^{-4}$ . Inflation washes out everything.

<sup>11</sup>General expression is  $\rho = \rho_\phi = \frac{1}{2} \dot{\phi}^2 + V(\phi) + \frac{1}{2a^2} (\vec{\nabla} \phi)^2$ ,  $p = p_\phi = \frac{1}{2} \dot{\phi}^2 - V(\phi) - \frac{1}{2a^2} (\vec{\nabla} \phi)^2$ .

We will assume the slow-roll approximation (for the flatness of the potential, the small kinetic term, and neglecting  $\ddot{\phi}$  in the equation of motion for the inflaton):

$$\epsilon \ll 1, \quad |\eta_s| \ll 1, \quad (2.37)$$

with the slow-roll parameters<sup>12</sup>

$$\epsilon \equiv -\frac{\dot{H}}{H^2}, \quad \eta_s \equiv \frac{\dot{\epsilon}}{H\epsilon}. \quad (2.38)$$

### Fluctuations in the inflaton field

We can now add small perturbations to the inflaton field, which will be the origin of cosmological fluctuations, such as the CMB and galaxy clustering. We consider an independent primordial fluctuation generated at some mode scale  $\lambda \sim k^{-1}$ , with wavenumber  $k$ , starting inside the comoving Hubble radius  $\mathcal{H}^{-1} = (aH)^{-1}$ .<sup>13</sup> During inflation the comoving Hubble radius decreases (e.g. [284]), so that the mode will exit the horizon at some  $k \sim aH_k$ , where subscript  $k$  denotes the moment of exit, only to reenter after inflation has ended at some other later time, and source late-time observables.

To understand scalar perturbations in inflation, we will consider a gauge with spatially flat slicing, so that the spatial part of the metric is  $g_{ij} = a^2\delta_{ij}$ . In this way, we can neglect metric perturbations, and consider only the ones for the inflaton (e.g. [81]).<sup>14</sup>

We can rewrite (2.35) using conformal time

$$\phi'' + 2\mathcal{H}\phi' - \nabla^2\phi + a^2\partial_\phi V = 0, \quad (2.39)$$

where  $\mathcal{H} = \frac{a'}{a}$ .

We can describe the inflation field as given by an unperturbed (classical) part  $\bar{\phi}(\eta)$  and a small (quantum) perturbation  $\delta\phi(\vec{x}, \eta)$  ( $\langle\delta\phi^2\rangle \ll \bar{\phi}^2(\eta)$ ), i.e.  $\phi(\vec{x}, \eta) = \bar{\phi}(\eta) + \delta\phi(\vec{x}, \eta)$ .

Plugging this into (2.39), and considering only the perturbations

$$\delta\phi'' + 2\mathcal{H}\delta\phi' - \nabla^2\delta\phi + a^2\partial_{\phi\phi}^2 V\delta\phi = 0. \quad (2.40)$$

<sup>12</sup>Note that  $\frac{\dot{a}}{a} = \dot{H} + H^2 = (1 - \epsilon)H^2$ , therefore for inflation  $\epsilon < 1$ .

<sup>13</sup>The reason for which we consider the comoving Hubble radius as the scale of reference is because we can write the comoving particle horizon as  $\int \frac{dt}{a} = \int \frac{da}{a} \frac{1}{aH} = \int d\ln a \frac{1}{\mathcal{H}}$ , and we can see that  $\mathcal{H}^{-1}$  determines if two regions separated by some comoving distance are going to communicate or not.

<sup>14</sup>In general perturbations in the metric and in the inflaton will be coupled, making calculations more tedious. To simplify, we use the concept of gauge-invariance, where in as long as we maintain the relevant degrees of freedom, we can choose our coordinate system to study the problem at hand.

We can now rescale  $\delta\phi$  as  $u \equiv a\delta\phi$ . Next, we turn to quantizing this rescaled function. Writing (2.40) in terms of  $u(\eta, \vec{x})$ , we obtain

$$u'' - \nabla^2 u + \left( a^2 \partial_{\phi\phi}^2 V - \frac{a''}{a} \right) u = 0 . \quad (2.41)$$

We see that  $\frac{a''}{a} = 2a^2 H^2 \gg a^2 \partial_{\phi\phi}^2 V$  from (2.37) and (2.38), and so we can neglect this  $a^2 \partial_{\phi\phi}^2 V$  term. Going to Fourier-space

$$u''_{\vec{k}} + \left( |\vec{k}|^2 - \frac{a''}{a} \right) u_{\vec{k}} = 0 . \quad (2.42)$$

We now proceed to quantization. Note that the equations of motion that we derived for  $u$  can be obtained from an action  $S = \int d\eta \mathcal{L} = \frac{1}{2} \int d\eta d^3 \vec{x} \left( u'^2 + \frac{a''}{a} u^2 - |\vec{\nabla} u|^2 \right)$ . We then quantize the theory by promoting  $u$  and the conjugate momentum  $\pi = \frac{\partial \mathcal{L}}{\partial u'} = u'$  to operators, using the bosonic equal-time canonical commutation relations

$$[\hat{u}(\eta, \vec{x}), \hat{\pi}(\eta, \vec{y})] = i\delta_D^{(3)}(\vec{x} - \vec{y}), \quad [\hat{u}(\eta, \vec{x}), \hat{u}(\eta, \vec{y})] = 0, \quad [\hat{\pi}(\eta, \vec{x}), \hat{\pi}(\eta, \vec{y})] = 0 . \quad (2.43)$$

We then expand the field operators in mode functions, making sure to have Hermiticity,

$$\hat{u}(\eta, \vec{x}) = \int \frac{d^3 \vec{k}}{(2\pi)^3} \hat{u}_{\vec{k}}(\eta) e^{i\vec{k}\cdot\vec{x}}, \quad \hat{u}_{\vec{k}}(\eta) = \hat{a}_{\vec{k}} u_k(\eta) + \hat{a}_{-\vec{k}}^\dagger u_k^*(\eta) \quad (2.44)$$

with the mode functions  $u_k(\eta)$  solutions of the classical equations of motion (2.42), and  $\hat{a}_{\vec{k}}^\dagger, \hat{a}_{\vec{k}}$  the creation and annihilation operators respectively, satisfying the condition

$$[\hat{a}_{\vec{k}}, \hat{a}_{\vec{k}'}] = [\hat{a}_{\vec{k}}^\dagger, \hat{a}_{\vec{k}'}^\dagger] = 0, \quad [\hat{a}_{\vec{k}}, \hat{a}_{\vec{k}'}^\dagger] = (2\pi)^3 \delta^{(3)}(\vec{k} - \vec{k}') . \quad (2.45)$$

Inserting the mode function expansion in the first expression of (2.43) we obtain

$$[\hat{u}(\eta, \vec{x}), \hat{\pi}(\eta, \vec{y})] = \frac{1}{(2\pi)^3} \int d^3 \vec{k} i W(u_k, u_{k'}) e^{i\vec{k}\cdot(\vec{x}-\vec{y})} \quad (2.46)$$

where we have the Wronskian  $W$  given by  $W(u_k, u_{k'}) = u_k (u_{k'}^*)' - u_{k'}^* (u_k)' = i$ .

We now impose (2.42) onto the mode functions. We note that during inflation  $H \approx \text{const}$ ,  $a(t) \propto e^{Ht}$ , so that  $a(\eta) \approx -\frac{1}{H_k \eta}$  around the time that a mode of comoving scale  $k^{-1}$  exits the horizon, where in this case  $H_k$  is the Hubble parameter at horizon exit. Therefore (2.42) becomes for the mode functions

$$u_k'' + \left( k^2 - \frac{2}{\eta^2} \right) u_k = 0 . \quad (2.47)$$

The general solution to this equation is (e.g. [30])

$$u_k(\eta) = \frac{e^{-ik\eta}}{\sqrt{2k}} \left( 1 - \frac{i}{k\eta} \right), \quad (2.48)$$

where we have imposed Minkowski initial conditions and the Wronskian condition too, for  $\eta \rightarrow -\infty$ ,  $u_k = \frac{1}{\sqrt{2k}}e^{-ik\eta}$ : given a mode  $k$ , by considering scales well inside the (Hubble) horizon we have that perturbations oscillate just as in flat space, and well outside the (Hubble) horizon, the perturbations  $\delta\phi$  will be constant. Indeed, after horizon exit, we have that  $|\eta| \ll k^{-1}$ , and we can write  $u_k(\eta) \approx -\frac{i}{k\eta\sqrt{2k}}$ , so that  $\hat{u}_{\vec{k}}(\eta) = -\frac{i}{k\eta\sqrt{2k}} (\hat{a}_{\vec{k}} - \hat{a}_{-\vec{k}}^\dagger)$ .

### The primordial curvature perturbation

Outside the horizon, we can calculate the power spectrum of the inflaton fluctuations (defining the vacuum state by  $\hat{a}|0\rangle = 0$  and  $\langle 0|\hat{a}^\dagger = 0$ )

$$\langle 0|\delta\hat{\phi}_{\vec{k}}\delta\hat{\phi}_{\vec{k}'}|0\rangle \equiv (2\pi)^3 P_{\delta\phi}(k)\delta_D^{(3)}(\vec{k}-\vec{k}') \stackrel{\text{Super Horizon}}{=} \frac{\langle 0|\hat{a}_{\vec{k}}\hat{a}_{\vec{k}'}^\dagger|0\rangle}{2\eta^2 a^2 k^3} = (2\pi)^3 \frac{H_k^2}{2k^3} \delta_D^{(3)}(\vec{k}-\vec{k}'), \quad (2.49)$$

where we used  $\eta = -1/(aH)$  during inflation, and from the commutator relation  $\hat{a}_{\vec{k}}\hat{a}_{\vec{k}'}^\dagger = \hat{a}_{\vec{k}}^\dagger\hat{a}_{\vec{k}'} + (2\pi)^3\delta_D^{(3)}(\vec{k}-\vec{k}')$ , and we denote the exit from the horizon as Super Horizon. As  $H$  is nearly constant in slow-roll inflation, we obtain a nearly scale-invariant, dimensionless power spectrum for a light scalar field

$$\Delta_{\delta\phi}^2 \equiv k^3 P_{\delta\phi}/(2\pi^2) \approx \left( \frac{H_k}{2\pi} \right)^2 \quad (2.50)$$

as a consequence of the quantum fluctuations  $\delta\phi$ .

A common variable chosen to study scalar perturbations is the comoving curvature perturbation  $\mathcal{R}$ , a gauge-invariant quantity, whose expression in the spatially flat gauge  $\mathcal{R}$  is (e.g. [81])

$$\mathcal{R} = -\mathcal{H} \frac{\delta\phi}{\dot{\phi}}, \quad (2.51)$$

with  $'$  a derivative with respect to the conformal time.

Therefore, the power spectrum of  $\mathcal{R}$  at horizon exit is

$$\Delta_{\mathcal{R}}^2(k) = \left( \frac{H^2}{2\pi\dot{\phi}} \right)^2 \Big|_{k=aH}. \quad (2.52)$$

We can parametrize the power spectrum of inflationary scalar perturbations as a power law, with respect to some scale  $k_*$ , amplitude  $A_s$ , and spectral index  $n_s$ ,

$$\Delta_{\mathcal{R}}^2(k) = A_s(k_*) \left( \frac{k}{k_*} \right)^{n_s-1} . \quad (2.53)$$

as the logarithmic derivative, with respect to  $k$ , of its logarithm is first order in the slow-roll parameters. Under the slow-roll approximation, it is possible to show that (e.g. [30])

$$n_s - 1 = -2\epsilon - \eta . \quad (2.54)$$

In particular, a scale-invariant spectrum will have a zero derivative with respect to scale, and therefore  $n_s = 1$ . For a blue spectrum  $n_s > 1$ , and for a red one,  $n_s < 1$ .

The spectral index gives us direct information about the shape of the inflaton potential (as it is related to the potential itself and its derivatives).

Soon after the end of inflation, the scalar field  $\phi$  starts oscillating near the minimum of its potential, producing many elementary particles, in a relatively unknown period called reheating [284]. We do not know well the equations governing the perturbations at that time. So, how do we relate inflation to late-time observables? Luckily, the wavelength of the perturbations that concern us was outside the horizon well before this uncertain time, satisfying a conservation condition that allows us to connect the distant past to the more recent past.

There is a general theorem [283] which states that whatever the constituents of the universe are, and their classical equations, there is always at least one conserved scalar adiabatic mode, such that  $\mathcal{R}' = 0$  (in the limit of  $k \ll aH$ ). In particular, single-field inflation excites this adiabatic mode, as there is no other source of fluctuations, and different components satisfy the same equation [284]

$$\frac{\delta\rho_i}{\rho_i(1+w_i)} = 3\mathcal{R}\frac{H}{a} \int a(t')dt' , \quad (2.55)$$

where  $w_i = p_i/\rho_i$  sets the equation of state for the energy component  $i$  with energy density  $\rho_i$  and pressure  $p_i$ , and  $\delta\rho_i$  is the energy density fluctuation. For example, for

matter  $m$  and radiation  $r$ , we have the relation  $\frac{\delta\rho_m}{\rho_m} = \frac{3}{4} \frac{\delta\rho_r}{\rho_r}$ .<sup>1516</sup> The violation of this relation would rule out single-field models of inflation [145].

Super-horizon primordial perturbations with scale given by  $k$  re-enter the comoving horizon when  $k \geq (aH)^{-1}$ . The connection between these primordial and late-time scalar fluctuations can be summarized, assuming linear evolution, as [32]

$$\mathcal{O}(\vec{k}, \eta) = T_{\mathcal{O}}(k, \eta, \eta_e) \mathcal{R}(\vec{k}, \eta_e), \quad (2.56)$$

where  $\mathcal{O}$  is the late-time observable,  $\eta_e$  the time of horizon exit, and  $T_{\mathcal{O}}(k, \eta, \eta_e)$  a transfer function that accounts for the sub-horizon evolution of perturbations after horizon re-entry.

An example of an application of this equation is the relation between the matter power spectrum and the primordial scalar power spectrum [81]

$$\Delta_m^2(k, z) = \frac{9}{25} \frac{k^4}{\Omega_m^2 H_0^2} T_m^2(k) D_m^2(k, z) \Delta_{\mathcal{R}}^2(k). \quad (2.57)$$

In this case the matter transfer function is given by the product of  $T_m$  for connecting to primordial fluctuations and  $D_m$ , a function describing the sub-horizon growth of perturbations at late-times.

### Primordial non-Gaussianity

One of the remarkable properties of the free single-field inflation is that perturbations are described by a Gaussian distribution, similar to the ground state of a harmonic oscillator [249], with only small deviations from it [172].

It might be that inflation could generate higher-order correlation functions, beyond the power spectrum, potentially due to interactions in the early universe [53].

<sup>15</sup>A simpler way to discuss this is by modelling the inflation  $\phi$  field as a clock, and its perturbation  $\delta\phi$  as a perturbation in the local time  $\delta\eta$ . Adiabatic fluctuations have the property that the perturbed state of the fluid at  $(\eta, \vec{x})$ , is the same as the one in a background universe at  $(\eta + \delta\eta(\vec{x}), \vec{x})$ . Then we have  $\delta\rho_i = \bar{\rho}_i(\eta + \delta\eta(\vec{x})) - \bar{\rho}_i = \bar{\rho}'_i \delta\eta$ , giving  $\delta\eta = \frac{\delta\rho_i}{\bar{\rho}'_i} = \frac{\delta\rho_j}{\bar{\rho}'_j}$ .

<sup>16</sup>Another way to achieve adiabatic conditions is thermal equilibrium between photons and baryons, giving naturally adiabatic conditions, leading to  $3 \frac{\delta T}{T} = \frac{\delta\rho_b}{\rho_b} = \frac{3}{4} \frac{\delta\rho_\nu}{\rho_\nu}$ , where  $\rho$  is the mass density for baryons, and  $T$  is the temperature (e.g. [284]). Similarly, even if dark matter does not interact with photons, they could have been in equilibrium in the past, as interaction rate increases with energy (depending on the specific interaction). A priori, there is no reason to expect adiabatic initial conditions, although they are not violated by any specific data [13]. With single-field inflation, however we expect adiabatic initial conditions even if dark matter and photons were not in thermal equilibrium.

A popular way to parametrize non-Gaussianity generated by some inflationary models is the primordial local non-Gaussianity

$$\Phi_{\text{NG}} = \Phi_{\text{G}} + f_{\text{NL}}[\Phi_{\text{G}}^2 - \langle \Phi_{\text{G}}^2 \rangle], \quad (2.58)$$

where  $\Phi_{\text{G}}$  is a Gaussian auxiliary field, and  $f_{\text{NL}}$  is a dimensionless non-linearity parameter (or in general some function). As this is the square of a Gaussian, this will generate a non-zero three-point function in the large-scale gravitational potential  $\Phi_{\text{NG}}$ .<sup>17</sup>

This concludes our overview of inflation. To summarize, we followed primordial fluctuations generated in the early universe until the end of inflation. Now, we study the evolution of perturbations after they re-enter the horizon, and propagate to the late-time observables. One of these is the CMB, the oldest light in the Universe.

## 2.4 How the Universe looks today

Our Universe is filled with cosmic microwave background radiation (CMB). Figure 2.2 is a picture of this radiation as seen by the Planck satellite experiment. From our measurements, we know the following:

- Today, it can be fitted by a black body spectrum with a temperature of  $T_{\text{CMB}} = 2.725\text{K}$  and brightness

$$B_{\nu}(T_{\text{CMB}}) = 2h\nu^3 \frac{1}{e^{h\nu/(k_{\text{B}}T_{\text{CMB}})} - 1}. \quad (2.59)$$

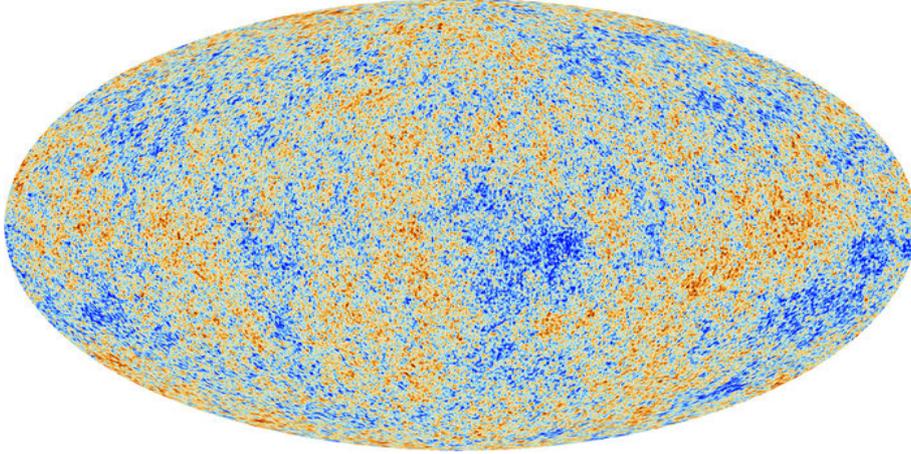
- There are approximately  $n_{\gamma} \sim 400 \text{ cm}^{-3}$  (photons per cubic centimeter).
- The CMB is an isotropic distribution over the sky to the precision of  $\frac{\delta T(\hat{n})}{T_{\text{CMB}}} \sim 10^{-5}$  (after removing the dipole anisotropy due to the motion of our Solar system with respect to the rest frame of the CMB).
- The CMB is weakly polarized, with radial/tangential patterns around temperature spots.

After years of understanding, we know that the CMB is the earliest picture of the Universe when it was only 380000 years old. It represents the free-streaming photons

---

<sup>17</sup>Note that the late-time non-Gaussianity we observe today in cosmological observables could come from different sources: the primordial non-Gaussianity we have just discussed, second-order non-Gaussianity arising from non-linearities (e.g. in the transfer function relating curvature perturbation to CMB anisotropies, or in gravitational evolution in the large-scale structure), secondary non-Gaussianity as given by gravitational lensing in the CMB or non-Gaussianity from foregrounds.

that were released when the universe was cool enough that electrons could combine with nuclei to form neutral hydrogen. The fluctuations shown in Figure 2.2 are a direct consequence of metric perturbations caused by primordial density fluctuations. Hence the CMB is a powerful probe of the physics of the early universe. In the next section, we will study these fluctuations.



*Fig. 2.2* Temperature fluctuations from the cosmic microwave background (CMB) that were mapped by the Planck Satellite experiment.

### 2.4.1 CMB fluctuations

The primordial fluctuations generated in the early Universe will affect metric perturbations, in turn perturbing the energy of CMB photons and leading to temperature anisotropies.

A basic quantity that we will focus on is the momentum of photons: as they propagate through space, photons will lose or gain momentum as they pass through potential wells and troughs. They will also experience a change in direction of the momentum due to gravitational deflection.

To study the evolution of the four-dimensional momentum of photons we will need the geodesic equation, and this requires specifying the metric.

Here we will consider only scalar perturbations to the metric, and we will take the space time metric in the Newtonian gauge

$$ds^2 = -a^2(\eta)(1 + 2\Psi(\vec{x}, \eta))d\eta^2 + a^2(\eta)(1 - 2\Phi(\vec{x}, \eta))\delta_{ij}dx^i dx^j, \quad (2.60)$$

where  $\Psi$  is Newton's gravitational potential, and  $\Phi$  is the spatial scalar curvature perturbation (deforms area, as it is a perturbation to the determinant of the spatial metric). If there is no anisotropic stress (e.g. from imperfect fluid)  $\Psi = \Phi$ , as can

be seen from Einstein equations in the Newtonian gauge. In this gauge, it is easier to interpret the formation of CMB anisotropies compared to other ones, as we can use Newtonian intuition for the gravitational redshift, and the infall of photons in potentials.<sup>18</sup>

Photons propagate in space along geodesics, as given by the geodesic equation for a freely falling body

$$\frac{d^2x^\mu}{d\lambda^2} + \Gamma_{\nu\rho}^\mu \frac{dx^\nu}{d\lambda} \frac{dx^\rho}{d\lambda} = 0, \quad (2.61)$$

where  $\lambda$  is an affine parameter along the photon's path. The second term in the equation keeps the form of the equation invariant under general coordinate transformations, a statement that gravity comes from geometry and is described by  $\Gamma_{\nu\rho}^\mu$ .

We study the evolution of the photon momentum  $p^\alpha = \frac{dx^\alpha}{d\lambda}$ , as it describes how photons lose or gain energy while propagating through our clumpy universe. It is one of the key ingredients to studying the CMB temperature anisotropies with the Boltzmann equation, which involves the photon distribution function  $f(x^\mu, p^\mu)$ .

In terms of the photon's momentum, the geodesic equation becomes

$$\frac{dp^\mu}{d\lambda} + \Gamma_{\nu\rho}^\mu p^\nu p^\rho = 0. \quad (2.62)$$

The photon 4-momentum  $p^\alpha$  is subject to the null geodesic condition  $g_{\mu\nu}p^\mu p^\nu = 0 = -a^2(1+2\Psi)(p^0)^2 + p^2 = 0$ , where we define  $p^2 = g_{ij}p^i p^j = a^2(1-2\Phi)\delta_{ij}p^i p^j$ . Then we have, when  $\Psi \ll 1$ ,

$$p^0 = \frac{p}{a\sqrt{1+2\Psi}} \approx \frac{p}{a}(1-\Psi). \quad (2.63)$$

We parametrize the photon momentum by the energy  $E$  measured by a comoving observer with four-velocity  $u^\mu$ , satisfying  $g_{\mu\nu}u^\mu u^\nu = -1$ . In an unperturbed FLRW universe, the comoving observer 4-velocity is  $u_\mu = -a\delta_\mu^0$  (by definition of the comoving frame). The measured energy is  $E = -p^\mu u_\mu = ap^0 = p$ , or  $p^0 = E/a$ , leading to  $p^\mu = \frac{E}{a}(1, \hat{p})$ , where  $\hat{p}$  is the photon propagation direction (normalized such that  $\delta_{ij}\hat{p}^i \hat{p}^j = 1$ ).

For a perturbed Universe, the same calculations as before lead to  $u_0 = -a(1+\Psi)$ , and  $E = -p^\mu u_\mu = a(1+\Psi)p^0$ . This appearance of  $\Psi$  for an observer at rest represents gravitational redshift. This will be important later for photons of the CMB.

---

<sup>18</sup>While studying perturbations for the single inflaton field, the gauge with spatially flat slicing is cleaner than the Newtonian one.

If we write the spatial part of the momentum with its direction  $\hat{p}$  and an amplitude  $A$ , then  $p^i = A\hat{p}^i$ . It follows that

$$p^2 = g_{ij}p^ip^j = a^2\delta_{ij}(1 - 2\Phi)\hat{p}^i\hat{p}^jA^2 = a^2(1 - 2\Phi)A^2. \quad (2.64)$$

Then, the amplitude of the spatial part of the momentum is  $A = \frac{p}{a\sqrt{1-2\Phi}} \approx \frac{p(1+\Phi)}{a}$ , so that when  $\Phi \ll 1$

$$p^i = A\hat{p}^i = p\frac{(1+\Phi)}{a}\hat{p}^i. \quad (2.65)$$

Then we finally obtain that the four-momentum can be written as

$$p^\mu = \frac{p}{a}(1 - \Psi, (1 + \Phi)\hat{p}^i). \quad (2.66)$$

From the 0 component of the photon's geodesic equation (2.62), and using the explicit expression for the Christoffel symbols (e.g. [21]), one can derive

$$\frac{1}{p} \frac{dp}{d\eta} = - \left( \mathcal{H} - \frac{\partial\Phi}{\partial\eta} + \hat{p}^i \frac{\partial\Psi}{\partial x^i} \right) = -\mathcal{H} - \frac{d\Psi}{d\eta} + \left( \frac{\partial\Psi}{\partial\eta} + \frac{\partial\Phi}{\partial\eta} \right). \quad (2.67)$$

This equation tells us how the photon energy changes as it propagates. The total derivative of the photon energy along its path is given by a series of terms in the middle side of this equation. First, a cosmological redshift (in absence of perturbations  $p \propto \frac{1}{a}$ ). We then have a term similar to a cosmological redshift: the spatial part of the metric has  $a^2(\eta)(1 - 2\Phi)d^2\vec{x}^2$ , so that photons locally experience a new scale factor  $\tilde{a}(\eta, \vec{x}) \approx a(\eta)(1 - \Phi)$ , and the photon momentum goes as  $p \propto \tilde{a}$ . Finally,  $\hat{p}^i \frac{\partial\Psi}{\partial x^i}$  represents gravitational redshift, and photons lose energy when climbing up a potential well. For example,  $\Psi < 0$  in a potential well, as a photon climbs up it  $\Psi$  increases and therefore  $\partial_i\Psi$  is positive, and the negative sign in the expression makes a photon lose energy.<sup>19</sup> The time dependence of the gravitational potentials is suppressed during matter domination, when potentials are constant on all scales and are contributing only at early-times during the matter-radiation transition, and at late-times when dark energy starts taking over. The latter effect will be important for example when studying dark energy with the CMB power spectrum.

Similarly, from the  $i$ th component of the geodesic equation, it is possible to derive an expression for the change in direction of the photons [54]

$$\frac{d\hat{p}^i}{d\eta} = -(\delta^{ij} - \hat{p}^i\hat{p}^j)\partial_j(\Psi + \Phi). \quad (2.68)$$

<sup>19</sup>There should also be a term representing gravitational waves changing in time, that we will not consider here for simplicity, and it is not dominant for the temperature anisotropies.

This term describes the gravitational lensing by gradients of scalar perturbations perpendicular to the line of sight. We will consider CMB lensing later in section (2.6), as it will be a second-order effect for our discussion here.

### The photon distribution

As the CMB is a blackbody to experimental accuracy, for the form of the photon distribution, we will consider the form of a Bose-Einstein distribution in equilibrium, where we describe the CMB with its temperature at position  $\vec{x}$ , with some direction  $\hat{p}$ , and frequency given by a momentum  $p$ :

$$f(\eta, \vec{x}, \hat{p}, p) = \left[ \exp \left( \frac{p}{\bar{T}(\eta)(1 + \Theta(\eta, \vec{x}, \hat{p}))} \right) - 1 \right]^{-1} \approx \bar{f}(p) \left[ 1 - \Theta(\eta, \vec{x}, \hat{p}) p \frac{\partial \ln \bar{f}}{\partial p} \right] \quad (2.69)$$

Here,  $\bar{f}(p, \eta)$  is the zero-order distribution function in absence of the temperature perturbation  $\Theta$ , and for the first-order correction part we have used

$$\bar{T} \frac{\partial \bar{f}}{\partial \bar{T}} = -p \frac{\partial \bar{f}}{\partial p}. \quad (2.70)$$

We parametrize the temperature fluctuations with  $\Theta$ , such that  $T = \bar{T}(\eta)[1 + \Theta(\vec{x}, \hat{p}, \eta)]$ . We assume here that the  $\Theta$  does not depend on  $p$ , i.e. there are no spectral distortions.<sup>20</sup>

We now want to turn to answering the question of how the temperature fluctuations evolve. To do this we must follow the photon distribution evolution.

## 2.4.2 Boltzmann equation

Our goal here is to describe the CMB anisotropies from primordial (curvature) perturbations.<sup>21</sup> With this goal in mind, there are two main requirements. First, to write out the evolution of the phase-space distribution of photons in the metric perturbed by the primordial density fluctuations, and second, to determine how these density fluctuations evolve. In this way, we will be able to obtain the energy density fluctu-

<sup>20</sup>However, spectral distortions could occur if free electrons have been heated significantly above the CMB temperature, or photons/energy has been dumped into the CMB from an external source, as it happens with secondaries in the CMB.

<sup>21</sup>As references for this section, we mainly follow [175, 21, 31, 144, 122].

ations of photons and relate these to the CMB anisotropies. In particular, in this section, we will derive the CMB temperature ones.<sup>22</sup>

First, we will focus on the (one-particle) phase-space distribution  $f(x^\mu, p^\mu)$ , a function of the space time position and four-momentum, that gives us the number of particles in an infinitesimal phase space volume  $d^3x d^3p$  as  $dN = f(x^\mu, p^\mu) d^3x d^3p$ . The evolution of the photon phase-space distribution is given by the Boltzmann equation [144]

$$\mathbf{L}[f] = \mathbf{C}_c[f] , \quad (2.71)$$

where  $\mathbf{L}$  is the Liouville operator and  $\mathbf{C}_c$  is a collision operator that contains all possible collision terms for  $f$ . It takes into account the observation that the (one-particle) phase-space distribution can change due to collisions which scatter a particle into or out of a volume element  $d^3x d^3p$  in phase space [85]. In its absence, the phase-space distribution is conserved along the photon's path, i.e.  $\mathbf{L}[f] = \frac{df}{d\lambda} = 0$ , with  $\lambda$  an affine parameter that parametrizes the photon's path.

Using the geodesic equation, it is useful to write explicitly the Boltzmann equation (we implicitly use the geodesic equation here):

$$\mathbf{L}[f] = \frac{df}{d\lambda} = p^\alpha \frac{\partial f}{\partial x^\alpha} - \Gamma_{\beta\gamma}^\alpha p^\beta p^\gamma \frac{\partial f}{\partial p^\alpha} = \mathbf{C}_c[f] . \quad (2.72)$$

We can see that the presence in the equation of the Christoffel symbol

$$\Gamma_{\beta\gamma}^\alpha \equiv \frac{1}{2} g^{\alpha\lambda} (\partial_\beta g_{\gamma\lambda} + \partial_\gamma g_{\lambda\beta} - \partial_\lambda g_{\beta\gamma}) \quad (2.73)$$

shows how the gravitational effects from the metric perturbations affect the distribution function.<sup>23</sup>

Looking at the 0th component of the photon momentum  $p^0 = \frac{dx^0}{d\lambda} = \frac{dx^0}{d\eta} \frac{d\eta}{d\lambda} \rightarrow \frac{d\lambda}{d\eta} = \frac{1}{p^0} \frac{dx^0}{d\eta} = \frac{1}{p^0}$ . We can then divide (2.72) by  $p^0$  and obtain

$$\frac{df}{d\eta} = \mathbf{C}[f] , \quad (2.74)$$

where the left-hand side is a total derivative with respect to the conformal time  $\eta$ , and we have defined the collision operator as  $\mathbf{C}[f] \equiv \frac{1}{p^0} \mathbf{C}_c[f]$ .

The left-hand side of the Boltzmann equation (2.74) can be expanded to give

---

<sup>22</sup>In this thesis we will mainly focus on the CMB temperature anisotropies. The CMB polarization theory for primordial anisotropies and weak lensing is omitted, although good reviews are in e.g. [131, 47, 153].

<sup>23</sup>Note, we assume the collision term local, so it will not be affected by the presence of gravity.

$$\frac{df}{d\eta} \equiv \frac{\partial f}{\partial \eta} + \frac{\partial f}{\partial x^i} \frac{dx^i}{d\eta} + \frac{\partial f}{\partial p} \frac{dp}{d\eta} + \frac{\partial f}{\partial \hat{p}^i} \frac{d\hat{p}^i}{d\eta} = \mathbf{C}[f] . \quad (2.75)$$

The comoving position time derivative can be calculated as

$$\frac{dx^i}{d\eta} = \frac{p^i}{p^0} = \hat{p}^i (1 + \Psi + \Phi) , \quad (2.76)$$

where we ignore second-order contributions in the potentials. We want then to substitute this expression in the Boltzmann equation. We note that the terms having  $\Psi + \Phi$  are negligible in the Boltzmann equation since they multiply  $\frac{\partial f}{\partial x^i}$  and this already contains first-order corrections, as the unperturbed distribution does not depend on the position.

We saw before that the  $\frac{d\hat{p}^i}{d\eta}$  term already depends on the gradient of the potentials in the direction perpendicular to the path of photons. Multiplying this term by  $\frac{\partial f}{\partial \hat{p}^i}$  gives a second-order contribution that we can ignore for now. We will return to this later when talking about CMB gravitational lensing in the late Universe.

Therefore, the Boltzmann equation simplifies to

$$\frac{\partial f}{\partial \eta} + \hat{p}^i \frac{\partial f}{\partial x^i} + \frac{\partial f}{\partial p} \frac{dp}{d\eta} = \mathbf{C}[f] . \quad (2.77)$$

Later we will address the collision distribution, and we will see that scattering starts at first order. Remembering the expression (2.69) we can write the previous one (at first order) as

$$-p \frac{\partial \bar{f}}{\partial p} \left[ \frac{d\Theta}{d\eta} - \frac{1}{p} \frac{dp}{d\eta} \right] = \mathbf{C}[f] . \quad (2.78)$$

Using equation (2.67) we are able to obtain to first order

$$-p \frac{\partial \bar{f}}{\partial p} \left[ \frac{\partial \Theta}{\partial \eta} + \hat{p}^i \frac{\partial \Theta}{\partial x^i} - \frac{\partial \Phi}{\partial \eta} + \hat{p}^i \frac{\partial \Psi}{\partial x^i} \right] = -p \frac{\partial \bar{f}}{\partial p} \left[ \frac{d\Theta}{d\eta} + \frac{d\Psi}{d\eta} - \frac{\partial(\Phi + \Psi)}{\partial \eta} \right] = \mathbf{C}[f] \quad (2.79)$$

If we set  $\mathbf{C}[f] = 0$ , the observed temperature perturbation is given by integration of this expression, from the time of the last scattering until today, along the line of sight. The presence of a potential well will redshift a photon, and the time variations in the scalar metric perturbations will boost photon energy if  $> 0$ , as happens during dark-energy domination, giving the late time-Integrated Sachs-Wolfe effect (late ISW). But what happens if we consider the interactions between photons and charged particles?

### Collision term

Before recombination, photons Thomson scatter efficiently off non-relativistic free electrons, as the rate of scattering is high compared to the Hubble expansion rate:

$$e^-(\vec{q}) + \gamma(\vec{p}) \longleftrightarrow e^-(\vec{q}') + \gamma(\vec{p}') . \quad (2.80)$$

We want to estimate changes in the photons ( $\gamma$ ) distribution related to Compton scattering.

In general, the collision term of the Boltzmann equation per unit proper time is given by [144]<sup>24</sup>

$$\begin{aligned} \hat{\mathbf{C}}[f] = & \frac{1}{E(\vec{p})} \int \frac{d^3\vec{q}}{(2\pi)^3 2E_e(\vec{q})} \int \frac{d^3\vec{q}'}{(2\pi)^3 2E_e(\vec{q}')} \int \frac{d^3\vec{p}'}{(2\pi)^3 2E(\vec{p}')} |\mathcal{M}|^2 (2\pi)^4 \delta_D^{(3)}(\vec{p} + \vec{q} - \vec{p}' - \vec{q}') \times \\ & \times \delta_D(E(\vec{p}) + E_e(\vec{q}) - E(\vec{p}') - E_e(\vec{q}')) [f_e(\vec{q})f(\vec{p}) - f_e(\vec{q}')f(\vec{p}')] , \quad (2.81) \end{aligned}$$

where  $f_e$  is the phase-space density of electrons,  $|\mathcal{M}|$  is the Lorentz invariant matrix element of the Compton scattering, averaged over initial and final spins,<sup>25</sup> and the delta functions impose energy and momentum conservation.

To lowest order in the electron energy transfer (in the CMB rest-frame), the following is true

$$|\mathcal{M}|^2 = 6\pi\sigma_T m_e^2 (1 + \cos^2\theta) , \quad \cos\theta = \hat{p} \cdot \hat{p}' . \quad (2.82)$$

For simplicity we ignore the  $\theta$  dependence in  $\mathcal{M}$ , and replace it by its angular average  $\frac{4}{3}$ , so that  $|\mathcal{M}|^2 = 8\pi\sigma_T m_e^2$ .

For non-relativistic Compton scattering, by performing the integrals in (2.81) it is possible to obtain (e.g. [81], [175])

$$\mathbf{C}[f] = -p \frac{\partial \bar{f}}{\partial p} \Gamma_T [\Theta_0 - \Theta(\hat{p}) + \hat{p} \cdot \vec{v}_b] , \quad (2.83)$$

where  $\mathbf{C}$  is calculated per unit conformal time,  $\Gamma_T = an_e\sigma_T$  is the scattering rate, given by the product of the Thomson-scattering cross-section  $\sigma_T$  and the electron number density  $n_e$ , and  $\vec{v}_b$  is the peculiar bulk velocity of the electrons. Finally,  $\Theta_0$  is the monopole of the temperature perturbation

<sup>24</sup>The factor of  $2E$  in the denominator is because  $\int d^3\vec{p} \int_0^\infty dE \delta_D(E^2 - p^2 - m^2) = \int d^3\vec{p} \frac{\delta_D(E - (p^2 + m^2)^{1/2})}{2E}$ . Also note the  $\frac{1}{E(\vec{p})}$ , is simply  $p$  for photons.

<sup>25</sup>We also implicitly assume invariance under time reversal, and that the density of electrons is low enough that we ignore the blocking emission factors that would appear in the collision term (e.g. if the final state is occupied by a fermion, a process should be blocked). See [144] for more details.

$$\Theta_0 = \int \frac{d^2\hat{p}}{4\pi} \Theta(\eta, \vec{x}, \hat{p}) . \quad (2.84)$$

The  $-\Theta$  term in (2.83) is related to photons scattering from a phase-space element  $(\eta, \vec{x}, p, \hat{p})$  to another one (reducing the distribution function), the term  $\Theta_0$  is related to photons with the arbitrary direction of propagation that scatter to align with  $\hat{p}$  (increasing the distribution function), and finally, the last term  $\hat{p} \cdot \vec{v}_b$  is related to the energy change of photons when they scatter off moving electrons (a Doppler effect). This is a very simple structure with only a non-vanishing monopole and dipole.

**Boltzmann equation for the photons** Now we have laid the foundation for studying the temperature anisotropies. Combining gravity effects on the perturbation in (2.79) with the scattering one in (2.83)

$$\frac{d\Theta}{d\eta} = \frac{\partial\Theta}{\partial\eta} + \hat{p}^i \frac{\partial\Theta}{\partial x^i} = + \frac{\partial\Phi}{\partial\eta} - \hat{p}^i \frac{\partial\Psi}{\partial x^i} + an_e\sigma_T [\Theta_0 - \Theta(\hat{p}) + \hat{p} \cdot \vec{v}_b] . \quad (2.85)$$

For convenience, we then define the optical depth,

$$\tau(\eta, \eta_0) = \int_{\eta}^{\eta_0} d\eta' \Gamma(\eta') = \int_{\eta}^{\eta_0} d\eta' a(\eta') n_e \sigma_T . \quad (2.86)$$

This defines the probability of no scattering for a photon traveling from  $\eta$  to  $\eta_0$  as  $e^{-\tau}$ . It is very large at an early time when the  $n_e$  is large, and very small at late times when  $n_e$  is small.

It is convenient to switch to Fourier space here as spatial derivatives will become factors of  $k_i$ , and as the perturbations are small we will have uncoupled differential equations for each Fourier mode<sup>26</sup>

$$\Theta(\eta, \vec{x}, \hat{p}) = \int \frac{d^3\vec{k}}{(2\pi)^3} e^{i\vec{k}\cdot\vec{x}} \Theta(\eta, k, \mu) \quad (2.87)$$

where  $\mu \equiv \hat{k} \cdot \hat{p}$ . We have (noting  $\frac{\partial\tau}{\partial\eta} = -n_e\sigma_T a$ )

$$\frac{\partial\Theta}{\partial\eta} + ik\mu\Theta = \frac{\partial\Phi}{\partial\eta} - ik\mu\Psi - \frac{\partial\tau}{\partial\eta} [\Theta_0 - \Theta + i\mu v_b] , \quad (2.88)$$

where we assume that the baryon velocity is irrotational  $\vec{v}_b \equiv i\hat{k}v_b$ .

Since the fluctuations from a single Fourier mode for scalar perturbations are axisymmetric (dependence only on  $\mu$ ), we can use the Legendre polynomials expansion

<sup>26</sup>Note the abuse of notation where the Fourier transform  $\tilde{\Theta}$  is replaced by  $\Theta$ , with the arguments distinguishing the two.

$$\Theta(\eta, \vec{k}, \mu) = \sum_{l=0}^{\infty} (2l+1)(-i)^l \Theta_l(\eta, \vec{k}) P_l(\mu), \quad (2.89)$$

where  $\Theta_l$  are multiple moments of the distribution

$$\Theta_l(\eta, k) = \frac{i^l}{2} \int_{-1}^1 d\mu \Theta(\eta, k, \mu) P_l(\mu). \quad (2.90)$$

We can use this expansion for equation (2.88) by multiplication by  $P_l$ , and integration over  $\mu$ , plus the recursion relation

$$\mu P_l(\mu) = \frac{l+1}{2l+1} P_{l+1}(\mu) + \frac{l}{2l+1} P_{l-1}(\mu) \quad (2.91)$$

for the term  $\mu\Theta$ , to obtain

$$\Theta'_l - k \frac{l}{2l+1} \Theta_{l-1} + k \frac{l+1}{2l+1} \Theta_{l+1} - \tau' \Theta_l = \delta_{l0}(\Phi' - \tau' \Theta_0) + \frac{1}{3} \delta_{l1}(k\Psi + \tau' v_b), \quad (2.92)$$

where we denote  $' \equiv \frac{\partial}{\partial \eta}$ , and use  $P_0 = 1$ ,  $P_1 = \mu$ .

We, therefore, obtain a system of coupled ordinary differential equations for  $\Theta_l(\eta, k)$ , where each multipole  $l$  is related to nearby ones. To close the system, we would also need to write the Boltzmann equation for dark matter and baryons, the evolution for the metric perturbations, and finally decide the initial conditions (e.g. [81]).

We notice that to know the multipole  $l$ , we need to solve for all the previous ones, and know the subsequent one. If we want to calculate the anisotropies to, say,  $l \in \mathcal{O}(10^3)$ , then we need to solve for more than  $10^3$  equations (or to the  $l_{\max}$  of choice in a real experiment, set by the experimental resolution for example).

A better way to solve the system is line-of-sight integration, developed in [248]. It helps to calculate the CMB multipoles faster than solving the coupled system, and it provides a simple way to understand the CMB power spectrum.

### Connecting to temperature anisotropies on a 2-D sky

Now we follow free-streaming photons from the last-scattering surface to the observation, as done in [248]. After rewriting the left-hand side of (2.88) with a total derivative, and multiplying by  $e^{-\tau+ik\mu\eta}$  and integrating over  $\eta$  from some initial time  $\eta_{\text{init}}$  to a receiving time  $\eta_0$ , it is possible to obtain (e.g. [21])

$$\Theta(\eta_0, \vec{k}, \hat{n}) = e^{ik\mu\eta} e^{-\tau} (\Theta + \Psi)|_{\eta_{\text{init}}}^{\eta_0} = \int_{\eta_{\text{init}}}^{\eta_0} d\eta' [\Psi' + \Phi' - \tau' \Psi - \tau' \Theta_0 - i\tau' \mu v_b] e^{-\tau(\eta, \eta_0)} e^{ik\mu\eta'} \quad (2.93)$$

where we use that  $\tau(\eta_0) = 0$ ,  $\Psi(\eta_0, \vec{x}_0)$  gives an unobservable monopole for the perturbations (and thus can be dropped as it is absorbed into  $T_{\text{CMB}}$ ), and for some early enough initial time  $\eta_{\text{init}}$  we can approximate  $\tau(\eta_{\text{init}}) \approx \infty$ . This means that any initial anisotropy is completely erased by Compton scattering.

Now we can write

$$\Theta(\eta_0) = \int_{\eta_{\text{init}}}^{\eta_0} d\eta e^{-ik\mu(\eta_0 - \eta)} S(\eta, \vec{k}) , \quad (2.94)$$

where we define the source term as

$$S(\eta, k) = (\Psi' + \Phi') e^{-\tau(\eta, \eta_0)} + g [\Theta_0 + \Psi + i\mu v_b] \quad (2.95)$$

and the visibility function  $g = -\tau' e^{-\tau}$ .

In the end, we get the Legendre coefficients by using the plane-wave expansion  $e^{-ik\mu(\eta_0 - \eta)} = \sum_l i^l (2l + 1) j_l(k(\eta - \eta_0)) P_l(\mu)$ . But before doing that we must account for the  $\mu$  dependence in the last term of  $S$ . To this goal, we note that  $S$  multiplies an exponential in equation 2.94. The  $\mu$  term in  $S$  can be obtained by taking the derivative  $\frac{1}{ik} \partial_\eta e^{ik\mu\eta}$ . By integrating by parts it is possible to turn the  $\mu$  part into a derivative operator and be able to obtain the Legendre coefficients (e.g. [81])

$$\Theta_l(\eta_0, k) = \int_{\eta_{\text{init}}}^{\eta_0} d\eta \tilde{S}(\eta, k) j_l(k(\eta - \eta_0)) . \quad (2.96)$$

where now we have defined

$$\tilde{S} = (\Psi' + \Phi') e^{-\tau(\eta, \eta_0)} + g \left[ \Theta_0 + \Psi - \frac{v'_b}{k} \right] - \frac{v_b}{k} \partial_\eta g . \quad (2.97)$$

Terms in the integral weighted by the visibility function  $g$  are localized near the last-scattering surface, while the ones proportional to  $e^{-\tau(\eta)}$  have an integrated effect till today.

**Instantaneous Recombination** A general physical understanding of the different terms in the line-of-sight solution can be achieved by noting that the visibility function is very peaked at around recombination, so we will approximate  $g(\eta) \approx \delta_D^{(1)}(\eta - \eta_{\text{rec}})$ , and the line-of-sight solution will give us several terms.

First,  $(\Theta_0 + \Psi)_{\text{rec}}$ , also known as Sachs-Wolfe effect (SW), due to radiation overdensities and photon redshifting due to potentials at emission. Then,  $\int_{\eta_{\text{rec}}}^{\eta_0} d\eta' (\Psi' + \Phi')$  also known as the Integrated Sachs-Wolfe effect (ISW), due to photons moving in evolving in time gravitational potentials. And finally,  $(i\mu k v_b)_{\text{rec}}$  is the Doppler effect, caused by the change in energy due to the photons scattering off moving electrons.

The solution for the multipoles  $\Theta_l$  will depend on  $\Theta_0$ ,  $\Psi$ ,  $\Phi$ ,  $v_b$ , and a few derivatives. The highest multipole entering these is  $\Theta_1$ , although as we ignored the angular dependence of the Compton scattering there is also a  $\Theta_2$  term. To have this accurately calculated we need to look at the first few multipoles of the temperature anisotropies in the Boltzmann hierarchy, up to some truncating  $l_{\text{max}} \in \mathcal{O}(10)$  [81].

**Acoustic oscillations** We will now study the tight-coupling limit ( $\Gamma \gg \mathcal{H}$ ), where the mean free path of the photons is much smaller than the horizon. It is possible to show ([81]) that in this case the only non-negligible moments  $\Theta_l$  are  $l = 0, 1$ , with all the other  $l$ s suppressed, giving

$$\Theta'_0 + k\Theta_1 = \Phi' , \quad (2.98)$$

$$\Theta'_1 - \frac{k}{3}\Theta_0 + \Gamma\Theta_1 = \frac{1}{3}(k\Psi + -\Gamma v_b) . \quad (2.99)$$

By considering the perturbed stress-energy tensor for photons ( $T_{\mu\nu} = \int \frac{d^3\vec{p}}{E(\vec{p})} f p_\mu p_\nu$ ) (e.g. [31]) it is possible to show that  $\Theta_0 = \frac{1}{4}\delta_\gamma$ , where  $\delta_\gamma$  is the density contrast of photons, and  $\Theta_1 \propto v_\gamma$ , the photon gas bulk velocity. Assuming that photons and baryons are tightly coupled via Compton scattering (until recombination),  $v_\gamma = v_b$ , and combining the solution for  $v_b$  from the baryon Euler equation with the photon Euler equation ((2.99) but with  $v_\gamma$  in it), and taking time derivatives for the photon continuity equation ((2.98) but with  $\delta_\gamma$  in it), it is possible to obtain (e.g. [21], [31])

$$\Theta''_0 + \frac{\mathcal{H}R}{1+R}\Theta'_0 + c_s^2 k^2 \Theta_0 = -\frac{1}{3}k^2\Psi + \Phi'' + \frac{\mathcal{H}R}{1+R}\Phi' , \quad (2.100)$$

where  $R = \frac{\rho_b}{\rho_\gamma + p_\gamma} = \frac{3}{4} \frac{\bar{\rho}_b}{\bar{\rho}_\gamma}$  is the baryon-to-photon ratio,  $c_s^2 = \frac{1}{3(1+R)}$  gives the sound speed, and as usual  $\mathcal{H} = aH$ . This equation represents the acoustic oscillations of the baryon-photon fluid: the left hand side has friction and pressure terms, and the right hand side gravitational and time dilation terms. The metric potentials are obtained from Einstein's equations, with contributions from dark matter. In particular, for constant metric perturbations (2.100) becomes the equation for a forced harmonic oscillator, and we get oscillations for  $\Theta_0$ , the source of the CMB peaks.

**Diffusion damping** We assumed that photons and baryons are tightly coupled, behaving as a single fluid. In reality, interactions between photons and baryons are not instantaneous, and become less efficient as we approach the epoch of decoupling: photons travel a finite distance between scattering, with a mean free path of Thomson scattering  $\lambda \sim \frac{1}{n_e \sigma_T}$ . The mean distance travelled by photons in a Hubble time is  $D = \frac{1}{\sqrt{n_e \sigma_T H}}$  [81], determining a diffusion scale  $k_D \sim 2\pi/D$ , projected onto the CMB as  $l_D \sim k_D d_A(t_{\text{dec}}) \sim 1500\text{--}2000$  around the photons' decoupling. The temperature perturbations get washed out by the diffusion of photons on scales  $l \geq l_D$ , resulting in a CMB power damped by a factor of  $e^{-l^2/l_D^2}$  [258, 81].

**Reionization** On their way to us, CMB photons are affected by a number of secondary effects. In particular, radiation from the first stars in the Universe reionizes the intergalactic gas, freeing electrons that scatter CMB photons, and reducing the amplitude of fluctuations on scales smaller than horizon size at reionization by a factor of  $e^{-\tau}$ , where  $\tau$  is the optical depth due to reionization. Later we will discuss other secondary effects on the CMB anisotropies.

### The CMB temperature power spectrum

The temperature fluctuations  $\Theta$  are given by a statistically isotropic field over a sphere (with directions  $\hat{n}$ ). This makes the use of a harmonic description of the fluctuations suitable as it is better than the real-space one in separating out different scales with different contributions (e.g. inflation affects all the scales, photon diffusion only small scales, etc.).

Second, CMB anisotropies are, with good approximation, linear in the primordial fluctuations, therefore if the primordial fluctuations are described by a complex Gaussian random field, the same will hold for the temperature anisotropies. We will hence need a mean, and a two-point function to describe them.

We employ the spherical harmonics  $Y_{lm}(\hat{n})$ , eigenfunctions of the Laplace operator on the sphere that form an orthonormal set of functions over the sphere

$$\int d^2\hat{n} Y_{lm}^*(\hat{n}) Y_{l'm'}(\hat{n}) = \delta_{ll'} \delta_{mm'} . \quad (2.101)$$

This allows us to write any scalar function on the sky as a sum of contributions over different angular scales with a multipole expansion. The integral is an angular integral done over the volume element  $d^2\hat{n} = d\Omega \equiv d\cos\theta d\phi$ , with  $\theta \in [0, \pi]$ ,  $\phi \in [0, 2\pi]$ .

We write the CMB temperature fluctuations as

$$\Theta(\hat{n}) = \frac{T(\hat{n}) - \bar{T}}{\bar{T}} = \sum_{l=0}^{\infty} \sum_{m=-l}^l \Theta_{lm} Y_{lm}(\hat{n}) , \quad (2.102)$$

such that the multipole coefficients are

$$\Theta_{lm} = \int d^2\hat{n} Y_{lm}^*(\hat{n}) \frac{T(\hat{n}) - \bar{T}}{\bar{T}} . \quad (2.103)$$

We will be more interested in the statistics of a CMB map rather than what the exact temperature is in a particular direction (and time). This is because, as initial conditions set up by inflation are stochastic, we cannot tell from theory what precise value the CMB temperature will take.

The  $\Theta$  are just fluctuations from an average temperature so that

$$\langle \Theta_{lm} \rangle = 0 , \quad (2.104)$$

and from statistical isotropy

$$\langle \Theta_{lm} \Theta_{l'm'}^* \rangle = \delta_{ll'} \delta_{mm'} C_l , \quad (2.105)$$

where  $C_l$  is the dimensionless temperature power spectrum.<sup>27</sup> Assuming Gaussian temperature fluctuations, the physics of the primordial CMB can be described by this quantity. Therefore in real experiments, we will want to estimate the temperature CMB power spectrum from CMB maps.

We can rewrite the temperature perturbations  $\Theta_l$  in terms of a transfer function  $T$  containing a line of sight integral over source term and the geometrical spherical Bessel functions  $j_l$ , (e.g. [81, 21]) giving an expression for the CMB power spectrum, assuming instantaneous recombination, (e.g. [21])

$$C_l = \frac{2}{\pi} \int \frac{dk}{k} \Delta_{\mathcal{R}}^2(k) T_l^2(k, \eta_*) . \quad (2.106)$$

This writing of this equation makes it clear how our basic model affects the observed power spectrum. If  $\Delta_{\mathcal{R}}^2(k)$  is scale-invariant we can see that the wiggles for  $l \geq 60$  in Figure 2.3 are due to the transfer function, that encodes information about the photon-baryon fluid and its interactions, and we can gain insight from just the CMB. On large scales,  $l \leq 60$ , the spectrum is almost flat (in  $l(l+1)C_l$ ): these scales correspond to scales larger than the horizon near recombination: therefore no causal physics could act on them, and they represent a direct contribution from the initial conditions, visible through the SW effect [191] (we neglect ISW effects here).

<sup>27</sup>In the literature it is also reported in  $\mu\text{K}^2$ , temperature squared units. To do this, one just multiplies by the average temperature of today.

Currently, we have access to only one realization of our Universe, and we cannot access the random process of creation to obtain a statistical average. The best that we can do for now is to construct an estimator for  $C_l$ . Assuming full sky, an estimator for the variance of the temperature fluctuations  $\Theta_{lm}$  is

$$\hat{C}_l = \frac{1}{2l+1} \sum_{m=-l}^l |\hat{\Theta}_{lm}|^2. \quad (2.107)$$

This has the property that

$$\langle \hat{C}_l \rangle = C_l \quad (2.108)$$

and, assuming Gaussian random variables,

$$\langle (\hat{C}_l - C_l)^2 \rangle = \frac{2}{N_{\text{modes}}} C_l^2 = \frac{2}{2l+1} C_l^2, \quad (2.109)$$

where  $N_{\text{modes}}$  is the number of modes at a given  $l$ .

The expected square difference, over realizations, between this power spectrum estimator and the real power spectrum, is non zero, and equal to the cosmic variance we already discussed in the beginning of this chapter. For higher and higher  $l$ s the relative difference becomes smaller and smaller, as we have access to a larger sample of the statistical realization of  $a_{lm}$  (with size given by  $2l+1$ , the number of modes) to estimate the power spectrum.<sup>28</sup> Cosmic variance then limits the accuracy of the CMB observation with theory, as seen in Figure 2.3 for the lowest multipoles.

Cosmic variance is a general property of the observables in our universe, and it does not hold only for the CMB. Despite this, we will see in Chapter 5 that it is possible to overcome the cosmic variance limitations by combining different observables of the same underlying fluctuations.

### Flat-Sky

For small angular separations on the sky, it is possible to take the limit of the spherical harmonic expansion to derive the temperature fluctuations in Fourier space [286]

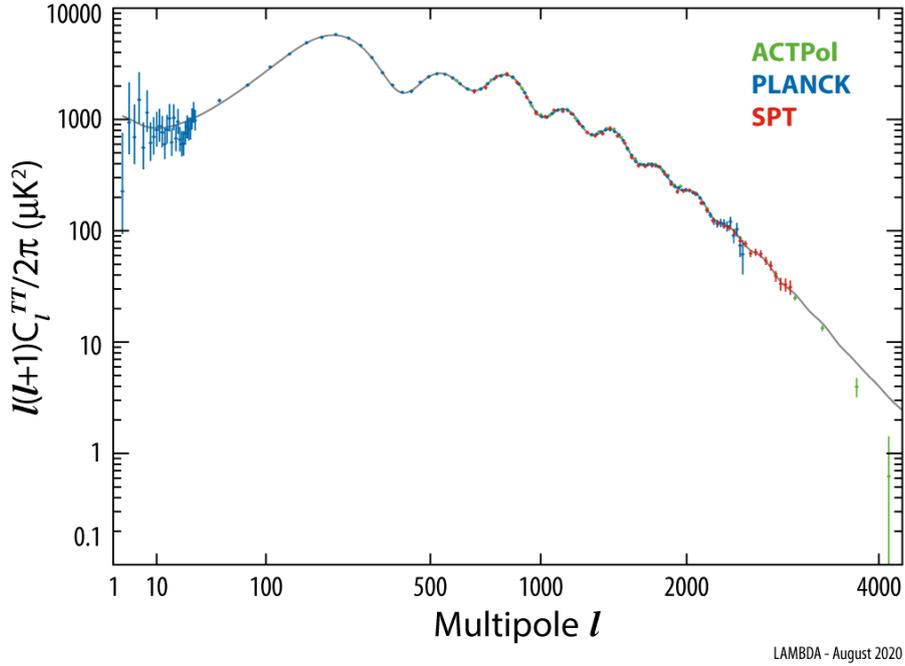
$$\delta T(\hat{n}) = \int \frac{d^2 \vec{l}}{(2\pi)^2} \Theta(\vec{l}) e^{i\vec{l} \cdot \hat{n}}, \quad (2.110)$$

and define the power spectrum as

<sup>28</sup>If we assume that the different  $|a_{lm}|^2$  are identically distributed and independent, then for smaller and smaller scales, the Gaussian approximation for the power spectrum holds. This in general is not true for the largest CMB scales ( $l < 30$ ). In these cases one might want to use other non-Gaussian approximations.

$$\langle \Theta(\vec{l}) \Theta^*(\vec{l}') \rangle = (2\pi)^2 \delta_D^{(2)}(\vec{l} - \vec{l}') C_l . \quad (2.111)$$

From now on, we will focus on the flat-sky approximation.



*Fig. 2.3* CMB (lensed) band powers from recent ACTPol [63], Planck [8], and SPT [269] measurements marginalized over galactic and extragalactic foreground emission and Sunyaev-Zel’dovich effects. In grey, a theoretical curve from parameters of a best-fit of Planck in  $\Lambda$ CDM. Image taken from LAMBDA <https://lambda.gsfc.nasa.gov/graphics/>.

### Cosmological parameters from the CMB

The shape of the CMB power spectrum is a direct probe of key parameters, such as the matter density  $\Omega_m h^2$ , baryon density  $\Omega_b h^2$ , and the shape of the primordial power spectrum. The overall amplitude is sensitive to  $A_s$ .<sup>29</sup> In particular, very large angular scales, larger than the sound horizon at decoupling, are not affected by oscillations in the photon-baryon fluid, and reflect the distribution of photons and potentials, directly connected to primordial quantum fluctuations.

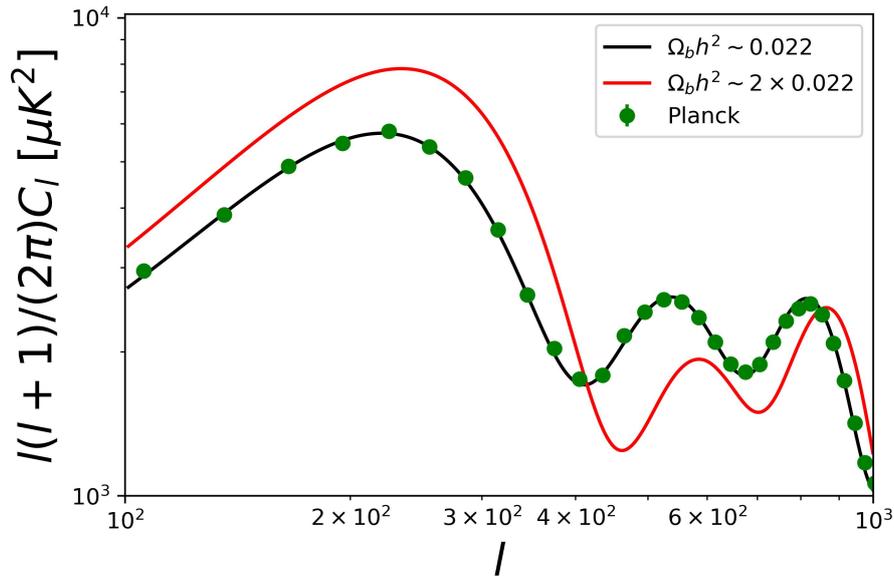
There are many ways to see how the CMB changes in practice, including:

- Constraining  $\Omega_b h^2$  from the CMB. The ratio of successive peak heights in the power spectrum is given by  $(1 + 6R)^2$  [31]: the larger the baryon-to-photon ratio

<sup>29</sup>Actually, the overall amplitude is sensitive to  $A_s e^{-2\tau}$ , where  $A_s$  is the amplitude of scalar primordial fluctuations, and  $\tau$  the optical depth to reionization (this degeneracy can be broken with polarization data, measuring the  $E$ -mode reionization bump on large scales).

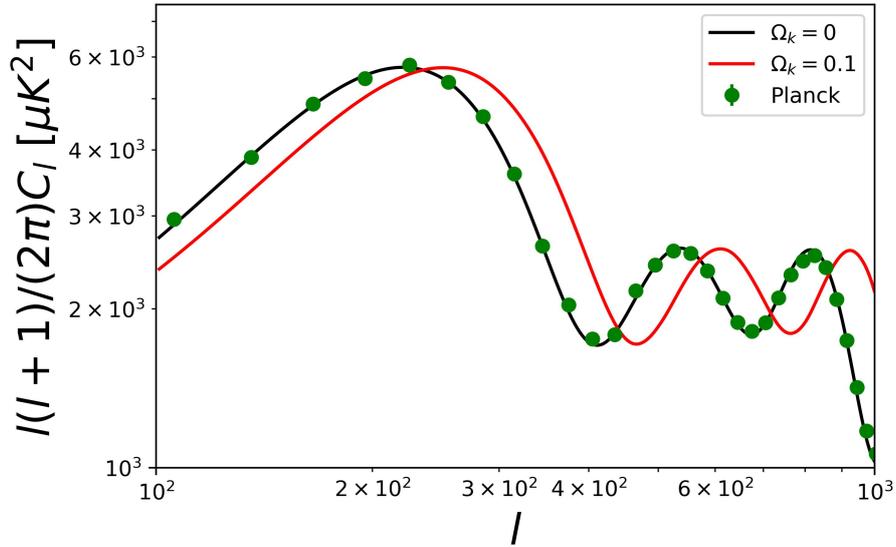
is, the bigger this parameter, making the compression in the baryon-photon plasma stronger (and the pressure will not be much affected by more baryons). As the baryons' physical density increases, therefore, the relative amplitude of odd acoustic peaks in the power spectrum is higher than the even-numbered ones, as can be seen from Figure 2.4. Also, adding more mass to the baryon-photon fluid will decrease the frequency of oscillations, as in a spring, pushing the position of the peaks to slightly higher multipoles  $l$ .<sup>30</sup>

- Constraining  $\Omega_k$  from the CMB. The geometry can be measured from the position of the first peak, as this depends on the projected size of the sound horizon at recombination,  $\theta_* \sim \frac{s_*}{d_A}$  where  $s_* \sim c_s \eta_*$ , depending on the angular diameter distance changes. We can see from Figure 2.5 that for  $\Omega_k > 0$ , the peaks are shifted to the right, meaning that a CMB fluctuation of a fixed intrinsic physical size will appear smaller compared to a flat universe.



*Fig. 2.4* CMB (lensed) power spectrum for a varying baryon density term, calculated from a best-fit cosmology given by Planck [8], using CAMB [154]. In black  $\Omega_b = 0.022$ , red  $\Omega_b = 2 \times 0.022$ , and in green Planck binned band powers as from LAMBDA [https://lambda.gsfc.nasa.gov/graphics/tt\\_spectrum/TT\\_data\\_2020aug\\_csv\\_format.dat](https://lambda.gsfc.nasa.gov/graphics/tt_spectrum/TT_data_2020aug_csv_format.dat). We can see the shift in acoustic peaks based on the baryon density term.

<sup>30</sup>The shape of the damping tail is sensitive to the physical baryon density too, allowing an independent determination. Therefore the damping tail of the power spectrum provides a consistency check for parameters extraction.



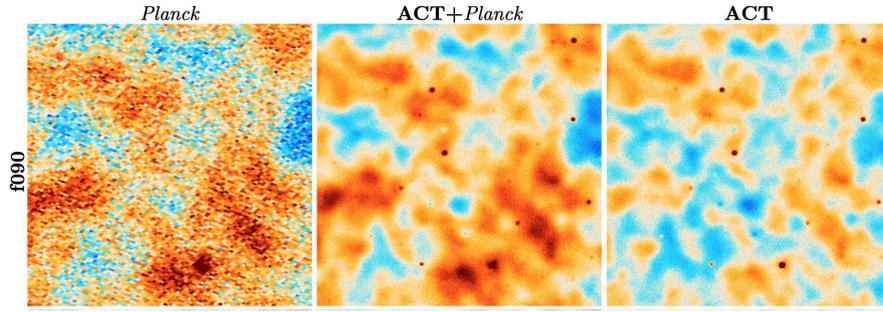
*Fig. 2.5* CMB (lensed) power spectrum for a varying curvature term, calculated from a best-fit cosmology given by Planck [8], using CAMB [154]. In black  $\Omega_k = 0$ , red  $\Omega_k = 0.1$ , and in green Planck binned band powers as from LAMBDA [https://lambda.gsfc.nasa.gov/graphics/tt\\_spectrum/TT\\_data\\_2020aug\\_csv\\_format.dat](https://lambda.gsfc.nasa.gov/graphics/tt_spectrum/TT_data_2020aug_csv_format.dat). We can see the shift in acoustic peaks based on the curvature term.

## 2.5 Secondary Anisotropies of the CMB

The primary CMB fluctuations are an unparalleled probe of the early Universe. Their primordial spectrum has three main components: on large scales, a plateau directly proportional to the primordial power spectrum; on intermediate scale, a series of peaks due to acoustic oscillations in the primordial photon-baryon fluid; and on small scales, a fall in power due to diffusion damping. Armed with this understanding, we explored fundamental cosmological parameters through these effects.

During their journey, CMB photons interact with the cosmic structure and experience gravity. An example is the (late) ISW effect, caused by time-varying potentials along the line of sight. Furthermore, our telescopes capture the CMB as well as light from other sources. We refer to all these effects as secondary anisotropies of the CMB.

In this section, we will discuss CMB foregrounds. Then, in the next one, the gravitational effect of lensing of the CMB photons, one of the main themes in this thesis. For reviews of secondaries of the CMB, see e.g. [126, 9].



*Fig. 2.6* Comparison between Planck only, ACT+ Planck ,and ACT only for a single frequency (90 GHz) in a small,  $3 \times 3 \text{ deg}^2$  patch of the sky. We can see how the higher resolution of ACT allows for detection of small-scale features, such as clusters, compared to Planck, which by contrast captures better the large scales compared to ACT alone. The figure is taken from [193], where there are many other examples that emphasize the difference between Planck and ACT on small scales.

### 2.5.1 Foregrounds

Foregrounds are all the radiation between us and the last-scattering surface contaminating the pure CMB. For example, galaxies and dust in our Universe contaminate CMB observations. Going to higher and higher resolution, CMB surveys become large-scale structure surveys. We can detect galaxy clusters by eye in a CMB map. Figure 2.6 shows this effect, where increasing resolution allows us to see, by eye, point sources or galaxy clusters.

Foregrounds are divided into two categories: galactic and extragalactic. Galactic foregrounds include thermal radiation of the interstellar medium dust, with an anisotropic distribution, or synchrotron radiation (for example see figures in [193]). Extragalactic foregrounds include light from background galaxies, and effects on the CMB due to scattering of CMB photons off gas in and around clusters and galaxies.

Both galactic and extragalactic foregrounds act as contaminants for CMB measurements: this is a problem for the interpretation of CMB observations when connecting to theory, as they bias the cosmological parameters. For example, foregrounds bias the inference of the spectral index  $n_s$  [185] , or the primordial local non-Gaussianity parameter  $f_{NL}$  [112].

A common solution, at the price of losing signal-to-noise, is to mask the most problematic areas of a map. People have also developed other methods to mitigate contamination of foregrounds in CMB maps.

In this section, we will only focus on extragalactic foregrounds, as these will be important for the main thesis topic of CMB lensing, as they dominate small angular scales, and we will discuss a very simple and blind method for foreground reduction.

The main idea of this method, as well as others, is to exploit the fact that the CMB and foregrounds have, in general, different emission laws, i.e. emission as a function of the frequency of observation. Then, we can construct maps of the sky where the CMB fluctuations are independent of frequency, while most of the foreground contamination is dependent.<sup>31</sup>

### tSZ effect

The thermal Sunayev-Zel'dovich [272] effect is a frequency-dependent change in the observed CMB temperature fluctuations due to the inverse Compton scattering of CMB photons off hot electrons along the line of sight, e.g. within galaxy clusters. This results in spectral distortions in the observed CMB, with fewer photons at lower frequencies, and more at higher frequencies.

The induced temperature fluctuation is given by an integral over the line-of-sight distance  $l$  of the electron pressure, for some direction  $\hat{n}$

$$\frac{\delta T_{\text{tSZ}}(\hat{n})}{T_{\text{CMB}}} = f_{\text{tSZ}}(\nu) \frac{\sigma_T}{m_e c^2} \int dl n_e(\hat{n}, l) k_B T_e(\hat{n}, l), \quad (2.112)$$

where (for the non-relativistic tSZ)

$$f_{\text{tSZ}} = x \frac{e^x + 1}{e^x - 1} - 4, \quad (2.113)$$

with  $x = \frac{h\nu}{k_B T_{\text{CMB}}}$ ,  $n_e$ ,  $T_e$  are the electron number density and temperature respectively,  $\sigma_T$  the Thomson cross-section,  $m_e c^2$  the electron rest-mass energy, and  $k_B$  the Boltzmann constant. For low frequencies/energies,  $f_{\text{tSZ}} = -2$ , resulting in a temperature dip.

As we expect to see ionized hot gas within galaxy clusters, the tSZ effect will be visible in the CMB in the direction of these, making it useful for detecting clusters, in particular high-redshift clusters, as the change in the CMB temperature fluctuations does not depend on the distance to the cluster (see e.g. [113]).

---

<sup>31</sup>Basically, the total sky brightness for several astrophysical components  $i$  is  $B(\nu, \hat{r}) = \sum_i B_i(\nu, \hat{r})$ , where  $\nu$  is the frequency, and  $\hat{r}$  unit vector the direction. The CMB contribution, for temperature across the sky  $T_{\text{CMB}}(\hat{r}) = \bar{T} + \delta T(\hat{r})$ , has an isotropic background component  $\bar{B}$  given by  $\bar{T}$ , and a correction  $B_{\text{CMB}}(\nu, \hat{r}) = \left(\frac{\partial \bar{B}(\nu)}{\partial T}\right) \delta T(\hat{r})$  (in general  $\delta T/\bar{T} \approx 10^{-5}$  for primordial fluctuations - we ignore dipole here, of order  $10^{-3}$ ). And so we can define a general temperature anisotropy for the emission  $i$  as  $\delta T_i(\hat{r}) = B_i(\nu, \hat{r}) / \left(\frac{\partial \bar{B}(\nu)}{\partial T}\right)$ .

### kSZ effect

The kinematic Sunyaev-Zel'dovich (kSZ) effect is a Doppler shift of the CMB due to Compton-scattering of CMB photons off free electrons of clusters with a non-zero line-of-sight bulk velocity (with respect to us). This induces a change in the observed CMB temperature given by (see e.g. [89])

$$\frac{\delta T_{\text{kSZ}}(\hat{n})}{T_{\text{CMB}}} = -\sigma_T \int dl e^{-\tau} n_e(\hat{n}, \eta) \vec{v}_e \cdot \hat{n} \quad (2.114)$$

where  $\tau$  the optical depth to Thomson scattering, and  $\vec{v}_e$  is the peculiar velocity of electrons.

The kSZ preserves the blackbody spectrum of the CMB, with a negligible small change in the blackbody temperature (to lowest order in  $(\frac{v}{c})$ , where  $v$  is the line-of-sight velocity), and is therefore difficult to distinguish from the CMB.

The kSZ effect arises not only in galaxy clusters but in general in ionized gas, such as during the epoch of reionization, when early quasars ionized the surrounding intergalactic medium.

### Cosmic infrared background

The cosmic infrared background (CIB) is composed of the cumulative infrared emission from dusty star-forming galaxies at different redshifts. Due to the spatial correlation of these galaxies, the CIB signal will be described by a clustered component, and a shot noise one, due to the discrete nature of sources assumed from a Poisson distribution. If galaxies trace the underlying matter distribution, then the CIB can be used to infer properties of the clustering of matter.

Formally speaking, the clustered CIB can be described by the halo-model ([67]), which assumes that galaxies reside in dark matter halos (according to some halo occupation distribution). Correlations can be large-scale ones, between two separate halos (2-halo term), or small-scale ones within a single halo (1-halo term). For a practical example of modeling the CIB, see [250], where the CIB power spectrum is described as an emissivity-weighted integral of the galaxy power spectrum. In contrast to the tSZ effect, the CIB is not a single field that can be rescaled across frequency channels according to a fixed spectral dependence. However, as an approximation, we can write the following spectral energy distribution for the CIB

$$f_{\text{CIB}}(\nu) \propto \frac{\nu^{3+\beta}}{e^{h\nu/(k_B T_{\text{CIB}})} - 1} \left( \frac{dB(\nu, T)}{dT} \Big|_{T=T_{\text{CMB}}} \right), \quad (2.115)$$

where  $\beta$  is a spectral index, and  $T_{CIB}$  is a dust temperature parameter.<sup>32</sup>

Foregrounds, with respect to the CMB, are a rich source of information about the Universe, and in recent years they have been a great source of focus not just for the purpose of understanding cosmological information, but also inferring information about galaxy formation (e.g. [18, 230]).

In the next section, we will see that, by treating foregrounds as contaminants to CMB observations and not as signals on their own, we can mitigate their effect on the measured CMB fluctuations.

### 2.5.2 A component separation method

If our goal is to study the CMB, then foregrounds can cause confusion in our understanding of the Universe, and it is thus important to remove these from our observations to extract the CMB.

There are multiple techniques to disentangle the CMB anisotropies from foregrounds, mainly based on the different spectral and spatial properties of the two. Some methods, known as non-blind methods, assume parametric forms for properties of the foregrounds. Then the parameters are fitted, given a likelihood, in a model that includes both CMB and foregrounds (e.g. [84]). One can then marginalize over the contributions of the foreground parameters.

Other methods, known as blind, assume only that the component of interest is dependent on the frequency of observation, in our case the CMB is known perfectly, with minimal assumptions about the foregrounds. A very popular and simple method is Internal Linear Combination (ILC), where observations of the sky at different frequencies are linearly combined to obtain an estimate of the CMB.

For a review of some popular foreground cleaning techniques see e.g. [76] (and for semi-blind methods see e.g. [280]). We now discuss briefly the ILC method.

We assume that the observed sky at  $N_f$  different frequencies can be written as

$$\vec{d} = \sum_{j \in \mathcal{S}} \vec{s}_j + \vec{n} , \quad (2.116)$$

where  $\vec{d}$  is the data vector whose components  $d_\nu$  are maps at a single frequency, and  $\vec{s}_j$  is the vector for the emission  $j \in \mathcal{S}$  (e.g. CMB, tSZ, ...), and  $\vec{n}$  is a noise component. As the ILC method is agnostic to the space of interest (e.g. real, harmonic, needlet), we will omit any dependence on some  $p$  in a pixel space, setting  $\vec{x} \equiv \vec{x}(p)$ .

---

<sup>32</sup>Note, this is only an approximate treatment for the CIB. In reality, as we have a sum of emissions from dusty galaxies, over a range of redshifts, we do not have one field with signal rescaled across frequencies. Different maps at different frequencies will include different CIB fluctuations. See for example [250], [177].

Given a component of interest  $\vec{c} \equiv \vec{s}_m$ ,  $m \in \mathcal{S}$ , we assume that the template of its emission is the same at all frequencies of observation,  $\vec{c} = \vec{a}_m s_m \equiv \vec{a} s$  where  $a$  is a recalibration coefficient and  $s$  is the template,<sup>33</sup> and that observations are calibrated with respect to the component of interest,  $\vec{a} \equiv \vec{e}$  where  $(\vec{e})_\nu = 1, \forall \nu$ . Therefore, we can write the problem as

$$\vec{d} = \vec{e}s + \sum_{j \in \mathcal{S}'} \vec{s}_j + \vec{n}, \quad (2.117)$$

where now  $\mathcal{S}'$  excludes the component of interest.

Now we form a linear combination of the observation at different frequencies with weights  $\vec{w}$ ,

$$\hat{s} = \vec{w}^T \vec{d}, \quad (2.118)$$

and with the condition that  $\vec{w}^T \vec{e} = 1$ , so that the linear combination is unbiased for extracting  $s$ .

In the ILC method, one seeks to minimize the variance of the above linear combination

$$\sigma^2 = \vec{w}^T R \vec{w}, \quad (2.119)$$

where  $R$  is the covariance matrix for the data. In particular, if the component of interest is uncorrelated with the other components (in this case the foregrounds, and the noise) then this means that we are seeking to minimize the total variance of foregrounds and noise.

Using Lagrange multipliers, it is possible to show that the following weights achieve this goal

$$\vec{w}_{\text{ILC}} = \frac{R^{-1} \vec{e}}{\vec{e}^T R^{-1} \vec{e}}, \quad (2.120)$$

where the denominator acts as the normalization to satisfy the condition  $\vec{w}^T \vec{e} = 1$ .<sup>34</sup>

There are two considerations for the ILC:

- If there are no foregrounds, then ILC leads to the minimum noise solution.
- If there is no noise, and some components are correlated, then to give the minimum variance output map, the ILC combination will create a biased map

<sup>33</sup>This is true for example for the CMB, kSZ, or (non-relativistic) tSZ, but not CIB, where the emission depends on the frequency too (e.g. [250, 177])

<sup>34</sup>This can be generalized to  $\vec{w}_{\text{ILC}} = \frac{R^{-1} \vec{e}}{\vec{a}^T R^{-1} \vec{a}}$  for recalibration coefficients  $a_\nu$ .

( $\hat{s} = s + b$ , where  $b$  will depend on the correlation of the components) with some foreground residuals remaining [76].

In general, even if in a true model the component of interest and other components (including noise) are uncorrelated, there might be some empirical correlation in the data covariance matrix. A large fluctuation in the data will result in a large fluctuation of the covariance matrix, as it depends on the data, leading to an excessively down-weighted mode in the ILC solution, and then leading to a bias. This will be particularly true for small data sets. There are ways to overcome this, mainly based on reducing this empirical correlation [77].

Another important point is if there is a second component, different from the one of interest but with the same frequency response, then this will propagate untouched to the final map. For our case, in general, an ILC combination to extract the primordial temperature anisotropies will be composed of the CMB+kSZ signal, as the latter has the same spectral signature as the former.

Given its simplicity, ILC is a very popular method with a few extensions: deprojecting unwanted emissions of the sky [224] by adding to the ILC problem a constraint of the form  $\vec{w}^T \vec{f}_j = 0$  for some foreground  $j$ , or by partially deprojecting the same component [2], or formally by optimizing the linear combination of data at different frequencies to account for foregrounds arising in CMB lensing [227].

## 2.6 CMB Weak Lensing

The CMB photons released at the last scattering surface will be deflected while propagating to us through the gravitational field generated by the large scale structure fluctuations. As a consequence, the observed CMB anisotropies are distorted, as are their statistics, and the angular power spectrum is modified with respect to the primordial CMB. It is therefore of great importance, for a complete understanding of observations, to account for this effect during a CMB cosmological data analysis. In addition, given the theoretical understanding of the primordial CMB, the CMB lensing signal can be extracted and used as an independent probe for the (projected) matter distribution in the Universe.

CMB lensing is a small effect, with an RMS of the deflection of around 2 arcmin, although the deflections are coherent over scales of a few degrees, making CMB lensing an important effect [153]. The small size of the CMB lensing effect made it observationally challenging to detect in earlier CMB experiments like WMAP. Nevertheless, by cross-correlating WMAP data with other tracers of large-scale structure, it was possible to get the very first detections of gravitational lensing using CMB temperature

at a  $3\sigma$  level [117, 262, 115]. And the first detection of CMB lensing from the CMB alone was from the Atacama Cosmology Telescope (ACT) collaboration [75].

In the next section, we will study the CMB lensing effect, then we will see how this affects the observed CMB.

## 2.6.1 The CMB lensing potential

### Weak CMB lensing deflection

By studying the term (2.68) it is possible to calculate the deflection angle that CMB photons experience while traveling from the last scattering surface (LSS) to us. The deflection angle will be the sum of several deflections, resulting in an integral over the photon path. Since deflections are small, we can make the assumption that the integral can be taken along the unperturbed photon's path, also known as the Born approximation [153].

We can write the deflection angle for direction  $\hat{n}$  as the gradient on the sphere of a projected CMB lensing potential<sup>35</sup>

$$\vec{d}(\hat{n}) = \vec{\nabla}_{\hat{n}}\phi(\hat{n}) , \quad (2.121)$$

where

$$\phi(\hat{n}) = -2 \int_0^{\chi_*} d\chi \frac{f_K(\chi_* - \chi)}{f_K(\chi_*)f_K(\chi)} \Psi_W(\chi\hat{n}, \eta_0 - \chi) , \quad (2.122)$$

where  $f_K(\chi)$  is the comoving angular diameter distance corresponding to the comoving distance  $\chi$  ( $= \chi$  for  $K = 0$ ),  $\chi_*$  is the comoving distance to last scattering,  $\eta_0$  is the conformal time at reception today, and  $\Psi_W = (\Psi + \Phi)/2$  the Weyl potential. From this expression we see that from  $f_K$  and  $\Psi_W$  we can probe the geometry and growth information of the Universe. Furthermore, as  $\Psi_W$  evolves with time, it affects the CMB temperature anisotropies through the ISW effect, with a main contribution from  $z \leq 2$  (when dark energy leads to decaying potentials). The same potentials gravitationally deflect CMB photons: there is a correlation between  $\phi$  and  $T$ . This is an effect important on large-scales, with little contribution on intermediate- and small-scales (see [153]), and so we will ignore it here.<sup>36</sup>

From now on we will assume a flat Universe, and for GR we can set  $\Psi_W = \Psi$  so that the expression for the lensing potential is

<sup>35</sup>Here we assume no curl component for the displacement field.

<sup>36</sup>Although it has important consequences for the estimation of some cosmological parameters, such as  $f_{NL}$ , as a bispectrum in  $T$  arises, see e.g. [143].

$$\phi(\hat{n}) = -2 \int_0^{\chi_*} d\chi W(\chi, \chi_*) \Psi(\chi \hat{n}, \eta_0 - \chi) , \quad (2.123)$$

where the CMB lensing kernel window function is

$$W^\phi(\chi) \equiv W(\chi, \chi_*) = \frac{\chi_* - \chi}{\chi_* \chi} . \quad (2.124)$$

To understand better the power of studying the CMB lensing potential, we will further assume linear evolution from primordial density fluctuations. Then considering  $\Psi$  in Fourier space, we have

$$\Psi(\vec{k}, \eta) = T_\Psi(k, \eta) \mathcal{R}(\vec{k}) , \quad (2.125)$$

where  $T_\Psi$  is the linear transfer function, and  $\mathcal{R}(\vec{k})$  is the primordial curvature perturbation, with a dimensionless power spectrum  $\Delta_{\mathcal{R}}(\vec{k})$ . Considering then the lensing potential in harmonic space [153], its power spectrum is:

$$C_L^{\phi\phi} = 16\pi \int \frac{dk}{k} \Delta_{\mathcal{R}}(k) \left[ \int_0^{\chi_*} d\chi T_\Psi(k, \eta_0 - \chi) j_L(k\chi) W(\chi, \chi_*) \right]^2 . \quad (2.126)$$

### Cosmological parameters from CMB lensing

If we are able to measure (2.126) we can infer cosmological parameters. To get a simple understanding of the dependence of  $C_L^{\phi\phi}$  on the cosmology, it is possible to derive an expression for the limit of high  $L > 100$ , or high redshift, around some fiducial cosmology for  $\Lambda$ CDM under several assumptions (see [203, 3]).<sup>37</sup> The basic dependence of the CMB lensing power spectrum on  $\Lambda$ CDM parameters can then be summarized as [3]

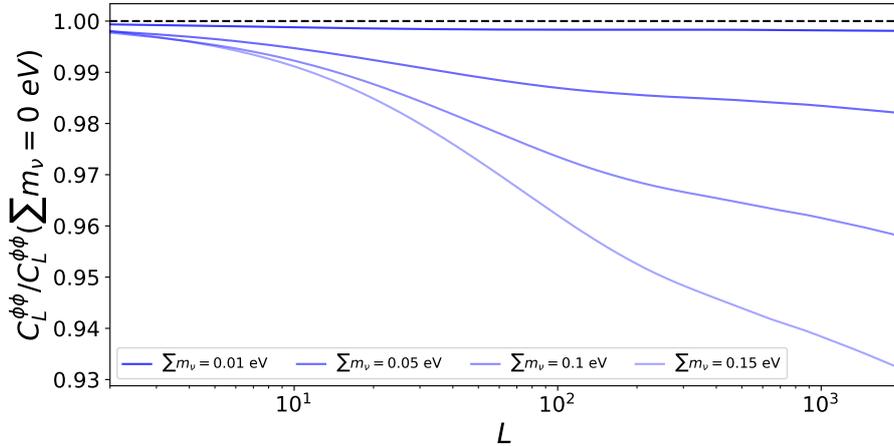
$$L^4 C_L^{\phi\phi} \propto A_s \Omega_m^{\alpha_L} h^{\beta_L} , \quad (2.127)$$

where  $\alpha_L, \beta_L$  are non-zero coefficients, in particular  $\alpha_L = 9/10$  and  $\beta_L = 3/2$  for  $L = 200$ .

Taking into account the full evolution of the matter power spectrum that enters the CMB lensing power spectrum other effects arise, e.g. due to massive neutrinos.

Massive neutrinos have important effects at the level of perturbations because of the free-streaming effect. Due to their large velocities, neutrinos do not cluster on small scales, but contribute to the background expansion of the universe. The overall effect is

<sup>37</sup>We assume the growth factor of perturbations to be one, a scale-invariant primordial power spectrum, and approximate the shape of the power spectrum near some fiducial model [3].



*Fig. 2.7* The effect of massive neutrinos on the CMB lensing potential as we increase  $\sum m_\nu$  from 0 meV to 150 meV. We fix other cosmological parameters in producing this plot, using CAMB [154].

that they suppress the growth of structure and so reduce the power spectrum of matter fluctuations (e.g. [86, 264, 140, 151]). This results in a scale-dependent suppression of  $C_L^{\phi\phi}$ , as seen in Figure 2.7. One can use CMB lensing to put an upper bound on the sum of neutrinos' masses, as shown for example in [256] where  $\sum m_\nu < 0.396$  eV (using ACTPol lens+ACTPol CMB + BAO, 95%). Tighter constraints are expected from future experiments, such as Simons Observatory (SO) with a forecast on the error bars of  $\sigma(\sum m_\nu) = 31$  meV (SO Goal+DESI-BAO) [5].

CMB lensing is also useful for breaking degeneracies between parameters constrained by the CMB alone. For example, the unlensed CMB has a geometric degeneracy, the angular diameter distance degeneracy. To explain this, we fix the comoving size of the sound horizon at decoupling  $s_*$ , at the fixed decoupling redshift  $z_{\text{dec}}$  (in essence, we fix the physical densities  $\Omega_b h^2, \Omega_c h^2$ ). Then, the angular scale of the primary acoustic peaks is [264, 85]<sup>38</sup>

$$l_a \sim \frac{D_*}{s_*}, \quad (2.128)$$

where  $D_*$  is the angular diameter distance to recombination. The quantity  $D_*$  depends on late-Universe parameters such as  $w, \Omega_K$  and  $\Lambda$ . But not all of them can be

<sup>38</sup>To be more precise, given a scale  $\lambda_n$  for the  $n$ th acoustic peak, the angle which subtends is given by:  $\theta_n = \lambda_n / D_*$ , with  $D_* = d_A(t_{\text{dec}})$  the angular diameter distance to the LSS. Roughly speaking, in spherical harmonics this angle corresponds to  $l_n \approx \pi / \theta_n = \pi D_* / \lambda_n$ . This can be seen in the simple case where we neglect the time variations in the potentials  $\Psi, \Phi$  in (2.100). Then one can obtain a harmonic oscillator equation with a constant gravitational forcing term, where there is a cosine  $\cos(k s_*)$  contributing to the acoustic oscillations, so we want  $k_n s_* = n\pi$ , with  $n$  constrained as an integer. Now,  $\lambda_n = 2\pi / k_n$ , where  $k_n$  is the wave-number of the peak, given by  $k_n s_* = n\pi$ . Therefore,  $l_n \sim \pi D_* / \lambda_n \sim \pi n \frac{D_*}{s_*} \rightarrow l_a = \frac{D_*}{s_*}$ .

constrained by  $D_*$ . Indeed, it is possible to vary any of the parameters in  $[w, \Omega_K]$ , by adjusting the dark energy density  $\Omega_\Lambda$  such that the angular diameter distance at recombination is unchanged, making the unlensed CMB power practically fixed.

The CMB lensing potential comes to the rescue. This can be written as a line-of-sight integral that includes geometric distances, and the power spectrum of the evolving potential. There is a dependence on late-time parameters such as  $[\Omega_\nu h^2, w, \Omega_K]$ , and varying these alters CMB lensing. Therefore we can break the angular diameter distance degeneracy with CMB alone, as shown in Figure 2.8 from [253].

Later, and in Chapter 3, we will discuss one of the most promising uses of CMB lensing: understanding the bias relation between the luminous observed tracers of matter, and matter itself. This is possible as CMB lensing is a direct probe of the projected matter distribution (via the Laplacian of the CMB lensing potential, with the connection made through the Poisson equation). Furthermore, the combination of CMB lensing with other probes of matter can play an important role in overcoming cosmic variance errors, allowing, in principle, much higher precision for parameters such as  $f_{\text{NL}}$ , or the sum of neutrino masses [237].<sup>39</sup>

We will now turn to the observable effects of lensing on the CMB temperature anisotropies, and later, given the richness of information that we can gain from the CMB lensing potential, how we it can be extracted from observations.

## 2.6.2 The lensed CMB

The observed lensed CMB temperature anisotropies  $T(\hat{n})$  are a remapping of the primordial unlensed CMB temperature anisotropies  $T^u(\hat{n})$  with a deflection angle  $\vec{d}(\hat{n})$ :<sup>40</sup>

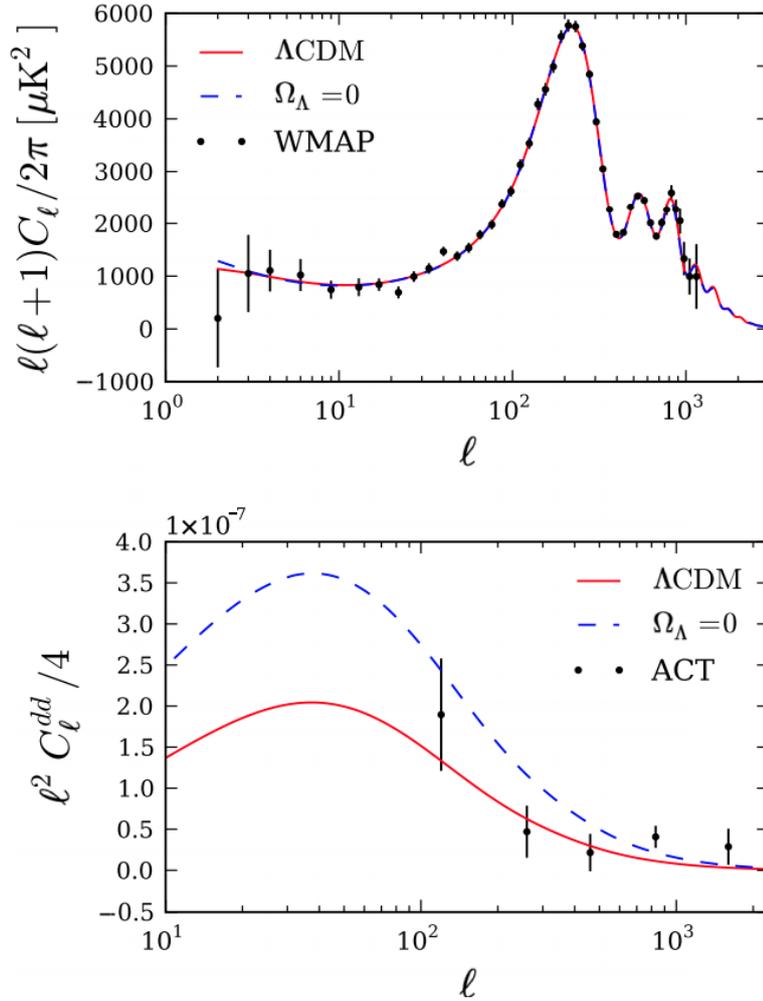
$$T(\hat{n}) = T^u(\hat{n} + \vec{d}(\hat{n})) . \quad (2.129)$$

If the deflection angle is small,  $|\vec{d}| \ll 1$ , we can express the temperature fluctuations as

$$T(\hat{n}) = T^u(\hat{n}) + \vec{d}(\hat{n}) \cdot \vec{\nabla} T^u(\hat{n}) + \frac{1}{2} \vec{d} \cdot \text{Hess}[T^u](\hat{n}) \vec{d}(\hat{n}) + O(|\vec{d}|^3) , \quad (2.130)$$

<sup>39</sup>Note there might be some factors that could degrade constraints when doing this type of analysis. For example, higher-order bias parameters could degrade constraints, if all free without priors. For other caveats see [237].

<sup>40</sup>A real observed CMB map is given by CMB anisotropies, plus a noise component due to finite experimental sensitivity, plus contamination as the SZ effects we discussed above. Here we have only the CMB.



*Fig. 2.8* Top panel: CMB  $TT$  power spectrum for two geometrically degenerate models: in blue, one with a curved Universe and no cosmological constant ( $\Omega_\Lambda = 0, \Omega_m = 1.29$ ); and in red a flat  $\Lambda$ CDM Universe ( $\Omega_\Lambda = 0.73, \Omega_m = 0.27$ ). The black dots are the seven-year WMAP temperature power spectrum data [147]. We can see that the data do not significantly favor either model. Lower panel: we show CMB lensing deflection power spectra for the same two models. We can see that they are no longer degenerate, with the  $\Omega_\Lambda = 0$  universe producing a lensing power spectrum larger than that measured by ACT data. Figure reproduced from [253].

where  $\text{Hess}[T^u]$  is the hessian of  $T^u$ , with components  $\text{Hess}[T^u]_{ij} = \nabla_i \nabla_j T^u$ .

Thanks to Helmholtz' theorem, the deflection field  $\vec{d}(\hat{n})$  can be decomposed into a curl-free part ( $\nabla \times \vec{\omega} = 0$ ), and a divergence-free one ( $\nabla \cdot \vec{v} = 0$ ). Neglecting linear-order tensor perturbations (e.g. from inflation), second-order scalar density fluctuations, and foreground contamination, we can ignore the curl mode (e.g. [194], [114]), and write the deflection angle as a pure gradient  $\vec{\nabla} \phi$ , where  $\vec{\nabla}$  is the covariant derivative on the unit sphere, and  $\phi$  a scalar field, the CMB lensing potential we discussed above. In Chapter ?? we will see how the curl mode can be use as a diagnostic to check for contaminants in the estimated gradient mode [66].

In general, as there are also low-redshift contributions to the CMB anisotropies coming from the large-scale structure, a CMB map would also include these secondary contributions,  $\tilde{X}_s$ . They will be correlated to the CMB lensing potential as it traces large-scale structure and a non-zero correlation with  $\tilde{X}_s$  will appear, inducing additional non-Gaussianities in the total observed CMB field (e.g. at the bispectrum level). We will ignore such biases for now, and return to them in Chapters 3, 4.

Focusing only on temperature anisotropies, in Fourier space, the expression above reads as

$$\begin{aligned} T(\vec{l}) &= \int d\hat{n} T(\hat{n}) e^{-i\vec{l} \cdot \hat{n}} \approx T^u(\vec{l}) + \delta T^u(\vec{l}) + \delta^2 T^u(\vec{l}) + \mathcal{O}(\phi^3) \\ &= T^u(\vec{l}) + \int \frac{d^2 \vec{l}'}{(2\pi)^2} T^u(\vec{l}') K^\phi(\vec{l}, \vec{l}') + \frac{1}{2} \int \frac{d^2 \vec{l}'}{(2\pi)^2} \int \frac{d^2 \vec{l}''}{(2\pi)^2} K^{\phi\phi}(\vec{l}, \vec{l}', \vec{l}'') T^u(\vec{l}') , \end{aligned} \quad (2.131)$$

where

$$K^\phi(\vec{l}, \vec{l}') = -(\vec{l} - \vec{l}') \cdot \vec{l}' \phi(\vec{l}') , \quad K^{\phi\phi}(\vec{l}, \vec{l}', \vec{l}'') = -(\vec{l}' \cdot \vec{l}'') [(\vec{l}'' + \vec{l}' - \vec{l}) \cdot \vec{l}'] \phi(\vec{l}') \phi^*(\vec{l}' + \vec{l}'' - \vec{l}) . \quad (2.132)$$

From this expression, we can now calculate the power spectrum for the lensed temperature CMB. The lensed power spectrum is given, assuming statistical homogeneity in Fourier space, by [129]:

$$\langle T(\vec{l}) T^*(\vec{l}') \rangle = (2\pi)^2 \delta_D^{(2)}(\vec{l} - \vec{l}') C_{|\vec{l}|} . \quad (2.133)$$

Substituting (2.131) and (2.132) into this, and assuming the CMB lensing potential is Gaussian and statistical independence between the unlensed temperature and the lensing potential, and using also  $\phi(\vec{l}) = \phi^*(-\vec{l})$ , the power spectrum of the lensed temperature CMB map is [123, 153]

$$C_{|\vec{l}|} = C_{|\vec{l}|}^u + C_{|\vec{l}|}^{\delta T^u \delta T^u} + C_{|\vec{l}|}^{T^u \delta^2 T^u} + \mathcal{O}(\phi^4) = \left[ 1 - \int \frac{d^2 \vec{l}'}{(2\pi)^2} C_{|\vec{l}'|}^{\phi\phi} (\vec{l}' \cdot \vec{l})^2 \right] C_{|\vec{l}|}^u + \int \frac{d^2 \vec{l}'}{(2\pi)^2} C_{|\vec{l}-\vec{l}'|}^{\phi\phi} C_{|\vec{l}'|}^u [(\vec{l}-\vec{l}') \cdot \vec{l}']^2. \quad (2.134)$$

The last term in (2.134) is a convolution term that blurs and smooths the power spectrum of the CMB peaks. The negative term in the big square brackets almost cancels in amplitude the convolution term, leaving the smoothing. We can see the overall effect in Figure 2.9, where we plot the fractional difference between lensed and unlensed CMB temperature power spectra, with respect to the unlensed CMB one.

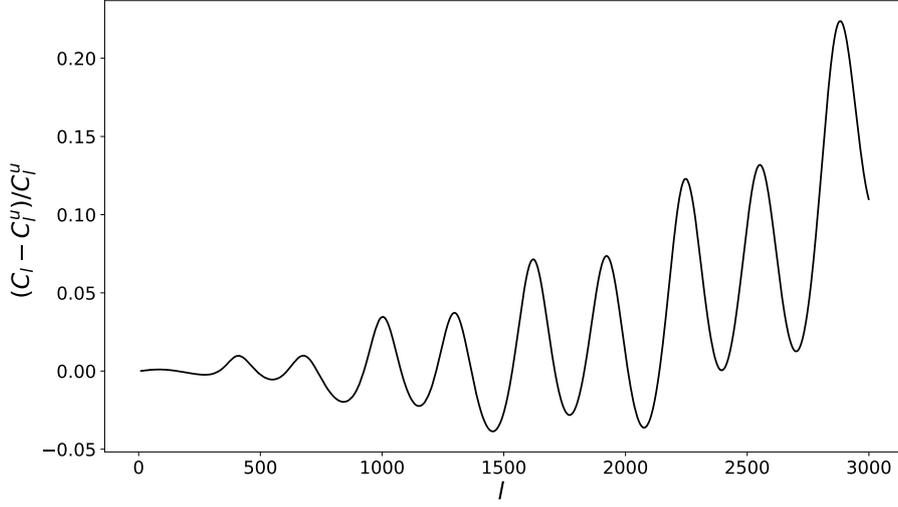
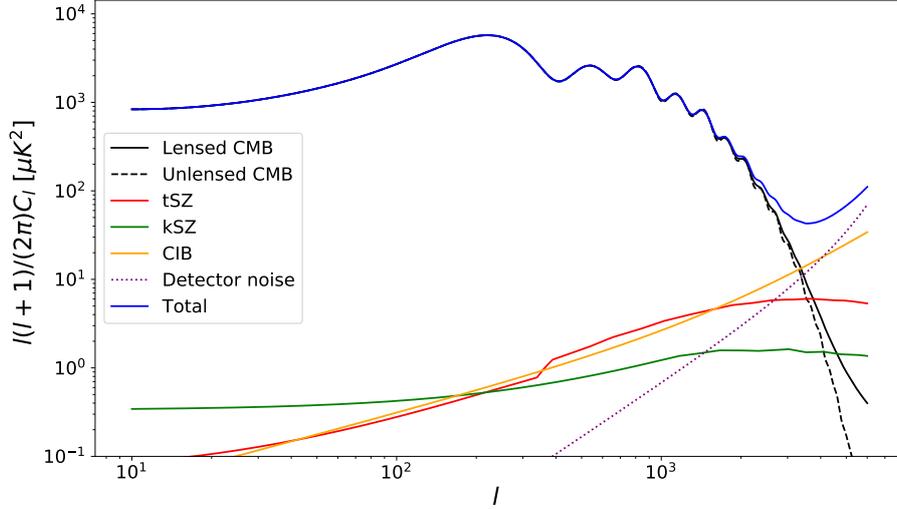


Fig. 2.9 The fractional difference between lensed and unlensed CMB temperature power spectra, with respect to unlensed CMB power spectrum.

The variance of the temperature field in real space is

$$\begin{aligned} \sigma^2(T) &= \int \frac{d^2 \vec{l}}{(2\pi)^2} C_{|\vec{l}|} = \\ &= \int \frac{d^2 \vec{l}}{(2\pi)^2} C_{|\vec{l}|}^u - \int \frac{d^2 \vec{l}}{(2\pi)^2} \int \frac{d^2 \vec{l}'}{(2\pi)^2} C_{|\vec{l}'|}^{\phi\phi} (\vec{l}' \cdot \vec{l})^2 C_{|\vec{l}|}^u + \int \frac{d^2 \vec{l}}{(2\pi)^2} \int \frac{d^2 \vec{l}'}{(2\pi)^2} C_{|\vec{l}-\vec{l}'|}^{\phi\phi} C_{|\vec{l}'|}^u [(\vec{l}-\vec{l}') \cdot \vec{l}']^2 = \\ &= \int \frac{d^2 \vec{l}}{(2\pi)^2} C_{|\vec{l}|}^u = \sigma^2(T^u). \quad (2.135) \end{aligned}$$

Weak CMB lensing conserves the total power of temperature fluctuations, redistributing it among different scales. In particular, for  $l \geq 2000$ , the CMB primordial unlensed



*Fig. 2.10* Example of power spectra of some of the sources discussed so far in the text. In solid black we have the lensed CMB, dashed black the unlensed CMB, in dotted purple some experimental Gaussian noise with white noise level of  $7\mu\text{Karcmin}$  and beam size of 1.4 arcmin. Then, we have foregrounds tSZ, CIB spectra at a single frequency of 148 GHz in red and orange, respectively, in green the kSZ signal, and finally in blue the sum of all the contributions. For this example, we use the szar code,<sup>41</sup> and access quickly to a precalculated theory CMB with the orphics code,<sup>42</sup> using spectra calculated from CAMB. The feature in the red line at around  $l \sim 300$  is from numerical fitting errors.

power spectrum is suppressed by diffusion damping, and therefore we expect some power transfer from large scales to small scales through the convolution between the lensing potential and the primordial CMB power spectrum. Going to very small scales,  $l \geq 3000$ , this can be clearly seen by the lensed temperature expansion  $T \sim \vec{\nabla}\phi \cdot \vec{\nabla}T^u$ , neglecting  $T^u$  due to diffusion damping. In (2.134) the convolution term will be mainly from  $|\vec{l}'| \ll |\vec{l}|$  for high  $|\vec{l}|$ , and the power spectrum will be simplified as  $C_l = l^2 C_l^{\phi\phi} \frac{1}{2} \langle |\vec{\nabla}T^u|^2 \rangle = l^2 C_l^{\phi\phi} \frac{1}{4\pi} \int \frac{dl'}{l'} l'^4 C_{|\vec{l}'|}^u$ , i.e. the lensed CMB temperature power becomes proportional to the unlensed gradient power and the deflection power  $l^2 C_l^{\phi\phi}$  [153].

Figure 2.10 presents a visual picture, where we see in solid black the CMB lensed power spectrum, and in dashed black the unlensed one, as well as the foregrounds components we already discussed for comparison.

We discussed the CMB lensing potential and its effects on the temperature power spectrum. What if we wanted to undo the effect of lensing, or wanted to use the potential as probe independent from the temperature fluctuations? The expansion  $T \sim T^u + \vec{\nabla}\phi \cdot \vec{\nabla}T^u$  shows that we might be able to recover  $\phi$  by  $\hat{\phi} \sim |T\vec{\nabla}T^u|$ , ignoring some spurious correlations for  $|T^u\vec{\nabla}T^u|$ . But before tackling CMB lensing

reconstruction, we will first discuss another tracer of matter fluctuations, then in section (2.8.1) we will discuss CMB lensing reconstruction.

## 2.7 Dark Matter and Biasing

Lensing is a powerful probe of the invisible (dark) matter field, although it gives only a direct projected mass distribution map of the Universe, i.e. two-dimensional information. In contrast, the matter field itself is three-dimensional, leading in principle to a higher number of modes that can be studied with respect to CMB lensing, although we cannot directly observe it.

Luckily, the distribution of galaxies in the Universe allows us to trace the three-dimensional matter field. But galaxies are biased with respect to the mass field, i.e. there is an excess of clustering of galaxies ( $g$ ) compared to the matter ( $m$ ) density field. One of the simplest ways to describe this, on large scales, is by measuring the auto/cross power spectra of the fields involved, and taking the time and scale-dependent ratios  $b_a = \sqrt{P_{gg}/P_{mm}}$ ,  $b_c = P_{gm}/P_{mm}$ .

To relate the underlying matter field to the observed galaxies we need an understanding of the bias. From the current standard cosmological theory, galaxies form in potential wells of virialized clumps of dark matter, also known as dark matter halos, where baryons can cool sufficiently fast [287]. Baryonic processes are very complicated to describe in detail, and simulations support our study of these phenomena. Nevertheless, it is still possible to gain understanding theoretically, with the common practice of separating galaxy clustering into two steps: first, non-linear evolution leads to clustering of dark matter halos, then these halos are occupied with galaxies (e.g. [22]).

Observationally speaking, we take the opposite route: we observe galaxies and study their statistics, and then we are able to make inferences about the non-linear matter field.

In the next section, we will start by focusing on the non-linear evolution of the matter field, with the goal of relating it to the initial linear (Gaussian) field. Then we will discuss galaxy clustering, with the goal of discussing the basics of biasing.

### 2.7.1 Late-time matter field

#### Newtonian Dynamics

In accordance with our current cosmological model, dark matter is the main component in which structure forms, so we will focus on it here. For the goals of this section, we

assume a pressureless, non-relativistic, collisionless dark matter fluid driven only by gravity.

During matter-domination, on sub-horizon scales, with small velocities, and for pressureless matter, the continuity and Einstein equations in the conformal Newtonian gauge correspond to the second-order Newtonian fluid equations: the continuity (or mass conservation) equation, Euler (momentum conservation) equation, and Poisson equation respectively (e.g. [21])

$$\delta' + \vec{\nabla} \cdot [(1 + \delta)\vec{v}] = 0 , \quad (2.136)$$

$$\vec{v}' + \mathcal{H}\vec{v} + (\vec{v} \cdot \vec{\nabla})\vec{v} = -\vec{\nabla}\Phi , \quad (2.137)$$

$$\nabla^2\Phi = \frac{3}{2}\Omega_m(\eta)\mathcal{H}^2\delta , \quad (2.138)$$

where the prime ' denotes the time derivative with respect to conformal time, and  $\vec{v}$  is the velocity perturbation, which can be decomposed in a curl-free part with  $\theta \equiv \vec{\nabla} \cdot \vec{v}$  and a divergence-free part.<sup>43</sup> We emphasize that for the Poisson equation at late times, energy density fluctuations are dominated by the fluctuations in the matter density, as we assume that the (dominant) dark energy component in the Universe is homogeneous. In addition, here we are considering, on top of a pressureless perfect fluid, no anisotropic stress.

Taking the Fourier transform of the continuity equation, and using the fact that  $\vec{v}(\vec{k}, \eta) = i\vec{k} \frac{\theta(\vec{k}, \eta)}{|\vec{k}|^2}$ , we have

$$\delta'(\vec{k}, \eta) + \theta(\vec{k}, \eta) = - \int \frac{d^3\vec{k}_1 d^3\vec{k}_2}{(2\pi)^3} \delta_D^{(3)}(\vec{k} - \vec{k}_1 - \vec{k}_2) \alpha(\vec{k}_1, \vec{k}_2) \theta(\vec{k}_1, \eta) \delta(\vec{k}_2, \eta) , \quad (2.139)$$

where we define the mode-coupling function

$$\alpha(\vec{k}_1, \vec{k}_2) = \frac{(\vec{k}_1 + \vec{k}_2) \cdot \vec{k}_1}{k_1^2} . \quad (2.140)$$

Similarly, the Fourier transform of the Euler equation, in conjunction with the Poisson equation (2.138), gives

---

<sup>43</sup>In absence of anisotropic stress, or primordial velocity with vorticity, we can neglect the vorticity term, as it decays at a linear level with the scale factor. This can be seen by solving the linear-order equation in the vorticity,  $\vec{\omega} + \mathcal{H}\vec{\omega} = 0$ , which comes from the linear order Euler equation separating scalar and vectorial parts, giving  $|\vec{\omega}| \propto a^{-1}$ , where  $\vec{\omega} = \vec{\nabla} \times \vec{v}$ . If there is no initial vorticity, and in absence of anisotropic stress, gravitational evolution will still not generate vorticity (although from simulations at late-times there is some evidence for it [21]).

$$\begin{aligned} \theta'(\vec{k}, \eta) + \mathcal{H}\theta(\vec{k}, \eta) + \frac{3}{2}\mathcal{H}^2\Omega_m(\eta)\delta(\vec{k}, \eta) = \\ = - \int \frac{d^3\vec{k}_1 d^3\vec{k}_2}{(2\pi)^3} \delta_D^{(3)}(\vec{k} - \vec{k}_1 - \vec{k}_2) \beta(\vec{k}, \vec{k}_1, \vec{k}_2) \theta(\vec{k}_1, \eta) \theta(\vec{k}_2, \eta) , \end{aligned} \quad (2.141)$$

where we have defined

$$\beta(\vec{k}, \vec{k}_1, \vec{k}_2) \equiv \beta(\vec{k}_1, \vec{k}_2) = \frac{(|\vec{k}_1 + \vec{k}_2|)^2 (\vec{k}_1 \cdot \vec{k}_2)}{2k_1^2 k_2^2} \quad (2.142)$$

representing the term  $(\vec{v} \cdot \vec{\nabla})\vec{v}$  in the Euler equation (2.137).

A closed-form solution for the system (2.139), and (2.141), in general, is not known. However, thinking about the large scales, where the fluctuations are small, linear theory is enough, and we can describe our fields with independent Fourier modes that have the same statistics as the primordial one. But when there are nonlinearities, there is a coupling of Fourier modes, and the analysis becomes more complicated. We can try to solve the above system under the assumption that we can expand the density and velocity fields perturbatively around linear solutions, using a power-law ansatz, in the limit of  $\delta \ll 1, \theta \ll 1$ . This is the basis of the next section, standard perturbation theory (e.g. [38]).<sup>44</sup>

### Standard Perturbation Theory

If we neglect all the quadratic terms in (2.139), with the limit of  $\delta \ll 1, \theta \ll 1$ , we obtain the following

$$\delta'(\vec{k}, \eta) + \theta(\vec{k}, \eta) = 0 . \quad (2.143)$$

Taking the conformal time derivative of this equation, substituting it in the linearized version of (2.141), and using the Poisson equation, we have the linear growth equation

$$\delta''(\vec{k}, \eta) + \mathcal{H}\delta'(\vec{k}, \eta) - \frac{3}{2}\mathcal{H}^2\Omega_m(\eta)\delta(\vec{k}, \eta) = 0 , . \quad (2.144)$$

This is a second-order differential equation with a damping term.

Seeing this in terms of the derivative with respect to the scale factor  $a$ , it is possible to solve (2.144) with growing and decaying-mode solutions, with separable spatial and

---

<sup>44</sup>In what follows, we will only discuss the Eulerian approach for the fluid equations, where we follow density and velocity fields of a fluid in a fixed coordinate system, as opposed to the Lagrangian approach, where we concentrate on the trajectory of individual particles in the fluid.

temporal parts,  $\delta(\vec{k}, \eta) = D_+(\eta)\delta_{+,0}(\vec{k}) + D_-(\eta)\delta_{-,0}(\vec{k})$ , where  $\delta_{+,0}(\vec{k})$  and  $\delta_{-,0}(\vec{k})$  are initial density perturbations.

The decaying-mode solution is (e.g. [21])

$$D_-(\eta) = D_{-,0}H = \mathcal{H}/a , \quad (2.145)$$

and the growing-mode solution

$$D_+(\eta) = D_{+,0}H \int_0^{a(\eta)} \frac{da'}{\mathcal{H}^3(a')} , \quad (2.146)$$

with  $D_{+,0}$  a normalization factor such that  $D_+(a=1) = 1$  (e.g. [21]). In a matter only Universe,  $H = a^{-3/2}$ , and thus  $D_+ = a$ ,  $D_- = a^{-3/2}$ .<sup>45</sup> As for structure formation, the decaying mode is not important, from now on, we will always focus on the growing-mode solutions,  $D \equiv D_+$ , or  $\delta_1(\vec{k}, \eta) = D(\eta)\delta_1(\vec{k})$ , with  $\delta_1(\vec{k})$  the linear density field today. Using this, and from the linearized continuity equation in Fourier space, and assuming a gradient only velocity, we can write the velocity divergence as  $\theta(\vec{k}, \eta) = -f(\eta)\mathcal{H}\delta_1(\vec{k}, \eta)$ , where we introduce the logarithmic growth rate  $f(\eta) = \frac{d \ln D}{d \ln a}$ . During matter domination this will be unity, and it will decrease at late times during dark energy domination.

To solve (2.139) and (2.141) at higher-order (with the non-linear terms) and on large enough scales to neglect vorticity effects, we expand the  $\delta$  and  $\theta$  in series of the  $n$ th-order solutions, and write the latter in powers of the linear solution  $\delta_1$  to obtain (e.g. [38, 242, 21])

$$\delta(\vec{k}, \eta) = \sum_{n=1}^{\infty} \delta^{(n)}(\vec{k}, \eta) = \sum_{n=1}^{\infty} \int \prod_{m=1}^n \frac{d^3 q_m}{(2\pi)^3} \delta_1(\vec{q}_m, \eta) F_n(\vec{q}_1, \dots, \vec{q}_n, \eta) (2\pi)^3 \delta_D^{(3)}(\vec{k} - \sum_{i=1}^n \vec{q}_i) , \quad (2.147)$$

and

$$\begin{aligned} \theta(\vec{k}, \eta) &= \sum_{n=1}^{\infty} \theta^{(n)}(\vec{k}, \eta) = \\ &= -f(\eta)\mathcal{H}(\eta) \sum_{n=1}^{\infty} \int \prod_{m=1}^n \frac{d^3 q_m}{(2\pi)^3} \delta_1(\vec{q}_m, \eta) G_n(\vec{q}_1, \dots, \vec{q}_n, \eta) (2\pi)^3 \delta_D^{(3)}(\vec{k} - \sum_{i=1}^n \vec{q}_i) , \end{aligned} \quad (2.148)$$

<sup>45</sup>For a  $\Lambda$ CDM Universe, there is a similar behavior deep in the matter domination phase, then the growth stalls as the cosmological constant starts to dominate.

where  $F_n, G_n$  are dimensionless, symmetrised kernels in their wavevector arguments. Focusing on an Einstein-de Sitter (EdS) cosmology, we can separate the time dependence of the density field as  $\delta^{(n)}(\vec{k}, \eta) = D^n(\eta)\delta^{(n)}(\vec{k})$ , where  $D = a$ .<sup>46</sup>

At first order  $F_1 = G_1 = 1$ , such that the expanded  $\delta$  is the linear field. At higher order, one substitutes (2.147) and (2.148) into (2.139) and (2.141), obtaining a recursion relation connecting the kernels  $F_n, G_n$  to lower-order ones  $F_m, G_m$ ,  $m < n$ , up to the more fundamental mode coupling functions  $\alpha$  and  $\beta$  (e.g. [21], [38, 242]).

As an example, the second-order solution for the density field gives a first non-trivial kernel, which will be important in Chapter 5, which can be obtained from (2.147) and (2.148) substituting into (2.139) and (2.141). Then we use first-order terms from the power series in the non-linear coupling terms on the right-hand side and second-order terms on the left hand side, plus the time dependence of the series ansatz, to obtain a coupled system from  $F_2, G_2$  that leads to

$$F_2(\vec{k}_1, \vec{k}_2) = \frac{5}{7} + \frac{\vec{k}_1 \cdot \vec{k}_2}{2} \left( \frac{1}{k_1^2} + \frac{1}{k_2^2} \right) + \frac{2}{7} \frac{(\vec{k}_1 \cdot \vec{k}_2)^2}{k_1^2 k_2^2}, \quad (2.149)$$

which can be written slightly differently as

$$F_2(\vec{k}_1, \vec{k}_2) = \frac{17}{21} + \frac{\vec{k}_1 \cdot \vec{k}_2}{2k_1 k_2} \left( \frac{k_2}{k_1} + \frac{k_1}{k_2} \right) + \frac{2}{7} \left[ \frac{(\vec{k}_1 \cdot \vec{k}_2)^2}{k_1^2 k_2^2} - \frac{1}{3} \right]. \quad (2.150)$$

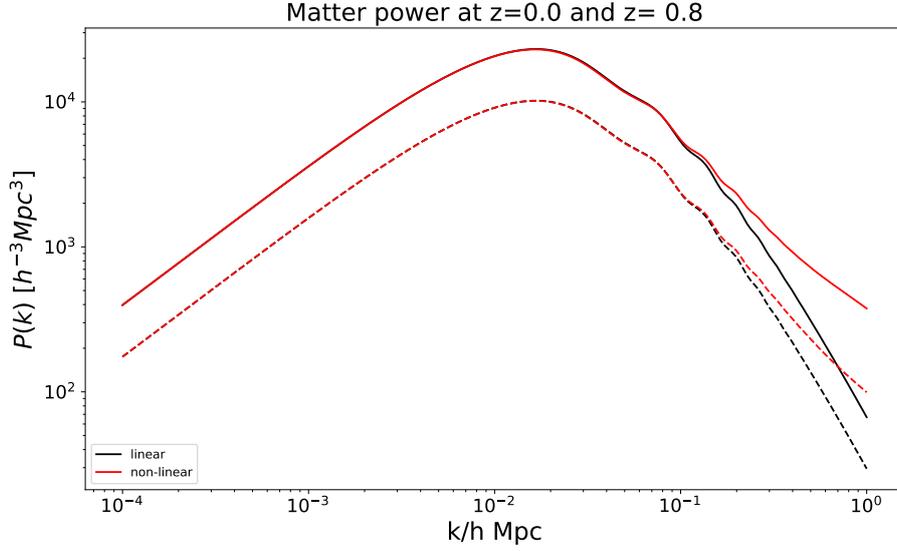
The terms in  $F_2$  correspond to the growth term, the advection term, and the tidal term respectively, (e.g. [252]). The growth term describes the enhanced growth factor in the presence of a long wavelength background mode, the advection term accounts for the displacement of initial perturbations due to bulk motions and the tidal term describes the mode coupling due to the tidal fields [21].<sup>47</sup>

Then, to obtain the non-linear matter power spectrum, or the  $n$ -point correlation functions, one plugs in the above series expansion, and using Wick's theorem the power spectrum is written as a sum of terms that include convolutions of powers of the linear matter power spectrum, with couplings given by the kernels. In Chapter 5 we will see this more explicitly.<sup>48</sup> In this way, one can connect the linear power spectrum to the process of seeding the fluctuations in the Universe, and hence extract information about the primordial Universe.

<sup>46</sup>Formally speaking, the equation should be  $\delta^{(n)}(\vec{k}, \eta) = D^n(\eta)\bar{\delta}^{(n)}(\vec{k})$ , where  $\bar{\delta}$  depends only on space. Although, unless ambiguous, we will just use  $\delta$ .

<sup>47</sup>There is another way to obtain this result. This is by noticing that a long-wavelength density perturbation corresponds to a tidal field on small scales. In the squeezed limit where the long-wavelength is much bigger than the short one on small scales, in an Einstein-de Sitter universe, it is possible to recover the standard perturbation theory kernel  $F_2$  [235].

<sup>48</sup>Here we assume a Gaussian linear matter field, such that contributions only come from even correlators.



*Fig. 2.11* Example of matter power spectrum. Solid lines are for  $z = 0$ , dashed ones for  $z = 0.8$ . Black is the linearly evolved power spectrum, while red accounts for non-linear corrections in the power spectrum. We can see how clustering changes going from dashed,  $z = 0.8$ , to solid  $z = 0$ , and how small scales are affected by non-linearity, with an enhanced power spectrum with respect to the linear case. Figure taken from <https://camb.readthedocs.io/en/latest/CAMBdemo.html>, using spectra calculated from CAMB.

Assuming an initial Gaussian linear random field that is also statistically homogeneous and isotropic, the evolved linear field will preserve the same properties, with a linear power spectrum  $P_1$

$$\langle \delta_1(\vec{k}_1, \eta) \delta_1(\vec{k}_2, \eta) \rangle = (2\pi)^3 \delta_D^{(3)}(\vec{k}_1 + \vec{k}_2) P_1(|\vec{k}_1|, \eta) . \quad (2.151)$$

In particular, by definition, the three-point function will be zero. Contrary to this, the evolved non-linear field  $\delta$  will acquire a bispectrum. Writing  $\delta \sim \delta^{(1)} + \delta^{(2)}$ , at leading order we have

$$\langle \delta(\vec{k}_1, \eta) \delta(\vec{k}_2, \eta) \delta(\vec{k}_3, \eta) \rangle = \langle \delta^{(2)}(\vec{k}_1) \delta^{(1)}(\vec{k}_2) \delta^{(1)}(\vec{k}_3) \rangle + 2 \text{ cyc.} , \quad (2.152)$$

leading to

$$B(\vec{k}_1, \vec{k}_2, \vec{k}_3) = 2F_2(\vec{k}_1, \vec{k}_2) P_1(k_1) P_1(k_2) + 2 \text{ cyc.} , \quad (2.153)$$

where cyc. is a cyclic permutation of the three vector modes. We can see that this term is only due to gravitational non-linearity. The shape dependence of this bispectrum is encoded in the  $F_2$  kernel. This observation is essential, as it makes the bispectrum a useful probe for cosmological parameters.

Standard perturbation theory is a great way to study the non-linear matter power spectrum. Despite this, it has some problems [21]:

- The power series expansion is not convergent, as on small scales the matter density is large.
- We assume a perfect pressureless fluid, when in reality, this is not valid on small scales, where shell crossing leads to multi-streaming.
- For certain initial power spectra, loop corrections to obtain the non-linear matter spectrum are UV divergent, requiring a UV cut-off that means the SPT predictions are all UV cut-off dependent, which is clearly unphysical.

To resolve these issues, today we have the Effective Field Theory of Large Scale Structure (EFTofLSS). The EFTofLSS provides answers to these shortcomings and provides physical, cut-off independent predictions. In this thesis, we will not use the EFTofLSS, although excellent reviews can be found elsewhere [33, 48].

During the matter-dominated era, the linear matter fluctuations grow as  $\delta \propto a$ , as long as  $\delta \ll 1$ . For  $\delta \gg 1$  this relation with the scale factor breaks for the non-linear matter density fluctuations, as these collapse gravitationally to form halos. Galaxies will be hosted by these halos. Dealing exactly with these non-linear processes is difficult, and usually, simulations or semi-analytic calculations are employed to study the formation and evolution of halos. However, we can treat the problem on large enough scales by describing dark matter halos and galaxies as biased tracers of the matter field.

## 2.7.2 Biasing

One of the main goals of this section is to show how we can extract information about fundamental physics by studying the observed luminous matter distribution as seen in galaxies. To this end, we want to relate the (mostly invisible to light) non-linear matter field to the galaxy field. In the current cosmological picture, galaxies form in potential wells of virialized clumps of dark matter, also known as halos, where baryons can cool sufficiently fast. Understanding the precise details of galaxy formation is a difficult task, that involves all the relevant processes of halo formation, as well as stochastic contributions due to the finite number of objects. Nevertheless, we can form a picture, on large enough scales, where we understand galaxy formation as follows: first, we have the clustering of dark matter halos, second, the occupation of these halos with galaxies. The result is that both halos and galaxies are seen as biased tracers of the underlying matter field.

Ideally, in our theoretical modeling, the galaxy bias should include non-local (e.g. containing gradients of the density field), and non-linear effects (e.g. higher-order biases), as for example represented by galaxy-formation physics. In Chapter 5 we will look in further detail at some of these terms. For now, our objective is to gain a basic understanding of biasing.

### Abundance of halos

A simple way to gain intuition about halo/galaxy bias is starting with the spherical collapse model, where regions in the (initially linear) density field exceeding a threshold  $\propto \delta_c$  today collapse to form bounded objects (see e.g. [22]). To determine the number density of collapsed objects of mass  $M$  at a given time, also known as the halo mass function, Press and Schechter (PS) [215] introduced the concept of smoothing the initial density field, with some filter of width  $R$  enclosing the mass  $M$ .

Suppose we generate the initial density fluctuations with a distribution  $P(\delta_0)$ . Then, the probability distribution for the linearly-evolved density fluctuations  $\delta_{\text{linear}}$  has the same form as the initial distribution, giving  $P(\delta_{\text{linear}})$  (e.g. [146]). The number of collapsed objects at a given time is related to  $P(\delta > \delta_c)$ , counting the fraction of the volume occupied by collapsed regions.

PS assumes a zero-mean Gaussian distribution, as current data suggests that the initial fluctuations are Gaussian or at least nearly Gaussian, i.e.

$$P(\delta_{\text{linear}}) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\delta_{\text{linear}}^2}{2\sigma^2}\right), \quad (2.154)$$

where  $\sigma^2$  is the variance of the density fluctuations.

Now we need to find overdense regions with a certain mass  $M$ . We use a sharp filter to bin the fluctuations. A filter  $W$  with width  $R$ , corresponding to a mass scale  $M = \frac{4\pi}{3}\bar{\rho}R^3$  ( $\bar{\rho}$  is the average matter density of the Universe),<sup>49</sup> gives a smoothed matter field, with original power spectrum  $P(q)$ , with a variance of

$$\sigma^2(M) = \int \frac{d^3\vec{q}}{(2\pi)^3} P(q) W^2[q, R(M)]. \quad (2.155)$$

We can see that  $\sigma^2$  is proportional to the power spectrum, and therefore its growth is given by the growth of matter density fluctuations squared,  $\sigma^2 \propto D^2$ . To trace

---

<sup>49</sup>Note that  $R$  is not the real radius of objects observed today with a given mass. It is rather the radius that these objects would have if they had a mean density equal to the mean mass density of the Universe. Regions then expand with the expansion of the Universe, turn around, and then contract to form objects with radii smaller than  $R$  (e.g. [146]).

this, a key parameter often used in cosmology is the variance of matter fluctuations smoothed on a scale of  $R = 8h^{-1}$  Mpc,

$$\sigma_8^2 \equiv \int \frac{d^3\vec{q}}{(2\pi)^3} P(q) W^2[q, R(M)] \quad (2.156)$$

From observations, its value is around  $\sigma_8 \approx 0.8$  for  $\Lambda$ CDM, and it is, in essence, a normalizing factor for the power spectrum.

Using a random walk in a Gaussian random field, PS derived that the number density of collapsed objects of mass  $M$  can be calculated as [215]

$$n(M) = -\sqrt{\frac{2}{\pi}} \frac{\bar{\rho}}{M^2} \nu \exp\left(-\frac{\nu^2}{2}\right) \frac{d \ln \sigma}{d \ln M}. \quad (2.157)$$

Actually, the mass function can be reduced to a (nearly) universal functional form for different redshifts and cosmologies, as tested in numerical simulations, when expressed in terms of the peak height <sup>50</sup>

$$\nu = \left(\frac{\delta_c}{\sigma(M)}\right)^2, \quad (2.158)$$

as

$$n(M) \equiv n(M, z) = \nu f(\nu) \frac{\bar{\rho}}{M^2} \frac{d \ln \nu}{d \ln M} \quad (2.159)$$

and  $f(\nu)$  is the fraction of mass that collapses into halos between  $\nu$  and  $\nu + d\nu$ .

Predicting the mass function  $n$  for different redshifts and cosmologies is crucial for interpreting observed cluster abundances. Despite the simple starting point of the spherical collapse model,<sup>51</sup> PS gives a good rough agreement with the mass function from  $N$ -body simulations, although it underestimates the abundance of massive halos. But measuring precisely the mass function is important, as it is an observable quantity that can predict  $\sigma^2$ , and from this the growth rate squared,  $D^2$ , as a function of redshift, to give constraints on models of dark energy, for example. To improve the agreement with simulations, there have been other proposals, such as the modified version of the PS mass function with fitting formulae to numerical simulations [257, 135], as well as non-Gaussian (in particular nearly Gaussian) extensions (e.g. see [23] and references therein).

<sup>50</sup>Definitions of  $\nu$  may vary in the literature, where some authors define it as  $\nu = \delta_c/\sigma$ . We follow the definition of [260].

<sup>51</sup>Plus an artificial factor of 2 in the mass function to account for the assumption that all the mass in the universe is enclosed in halos [215], that we did not discuss here.

## Biasing

A very simple way to understand halo biasing is with the Peak-Background Split argument (PBS) ([27, 65, 257]): halos in general form from short-wavelength fluctuations, as they have a larger root mean squared amplitude and are more likely to cross the critical collapse density;<sup>52</sup> the short-wavelength's fluctuations are modulated by the presence of long-wavelength fluctuations. A positive long-wavelength fluctuation will amplify the small scale one, making it easier to cross the critical threshold, or equivalently the collapse threshold goes from  $\delta_c$  to  $\delta_c - \delta_l$ , thus affecting the number density  $n$

$$n \rightarrow n_{\text{new}} = n - \frac{dn}{d\delta_c} \delta_l, \quad (2.160)$$

where  $\delta_l$  is the large-scale density mode.

This simply gives, to first order, a large-scale bias

$$b = \frac{\delta_g}{\delta_l} = -\frac{d \ln n}{d\delta_c}, \quad (2.161)$$

where we define  $\delta_g = n_{\text{new}}/n - 1$ .

On large scales, there is a linear relationship between the large-scale mode and the change in the number density of objects, so that the shape of the galaxy and matter power spectra are the same.

On small scales, we must use other models for the clustering of galaxies. A popular one is the halo model [67], built under the ansatz that all dark matter is partitioned into halos. Galaxies trace dark matter halos on large scales with a large scale bias, and on small scales the clustering is different from that of halos, counting correlations between galaxies in the same halo.

In general, we can approximate the relation between halos and the matter field with a local one:

$$\delta_g(\vec{x}) = B[\delta_m(\vec{x}), \epsilon], \quad (2.162)$$

where  $B$  is some functional, and  $\epsilon$  is a random variable representing the stochasticity of the galaxy bias, as the relation between the galaxy and matter field is not deterministic (e.g., [290]). The simplest form of stochasticity is white noise, uncorrelated with the matter field  $\langle \epsilon \delta \rangle = 0$ .

---

<sup>52</sup>This can be seen in our simple model by considering  $\sigma^2(M)$  for smaller and smaller radius, giving a bottom-up scenario of structure formation, starting from small objects that merge to form larger ones.

In Chapter 5 we will extend the definition of bias, although for now, it is interesting to look at one particular example ([92, 178])

$$\delta_g(\vec{x}) = b_1\delta_m(\vec{x}) + \frac{b_2}{2!}\delta_m^2(\vec{x}) + \frac{1}{3!}b_3\delta_m^3(\vec{x}) + \epsilon, \quad (2.163)$$

where  $b_i$  are bias parameters. Their precise value depends on the galaxy formation process, as well as on the physical properties of the galaxies. These could be, on their own, interesting, although usually when doing cosmological inference from data using our model, we do not care about the precise value of the biases, as we are only concerned about cosmological parameters. In absence of a more sophisticated theory, or faithful numerical simulations, in general, bias parameters will be marginalized over, at the price of increased error bars of the inferred cosmological parameters. To avoid a large inflation of the error bars, semi-analytic methods have been developed, combining fits to N-body simulations with perturbation theory calculations, in such a way that non-linearities in the matter fluctuations and non-linear bias terms can describe the galaxy field (e.g. [148, 281, 187, 56, 57]).<sup>53</sup>

## The Bispectrum

Until now, we have focused on the power spectrum of the matter field. This is the only quantity we need, if matter density fluctuations today are Gaussian. But gravitational non-linear evolution induces non-Gaussianities in the matter field today (i.e. non-zero connected  $n$ -point functions for  $n > 2$ ), making the power spectrum incomplete for a full understanding. Here we will focus on the large-scale bispectrum. This can be combined with the power spectrum to measure non-linear biases and break the degeneracy between the linear bias on large scales and the amplitude of the power spectrum.

A simple toy model is given by a local bias expansion, where the halo field is given in real space by

---

<sup>53</sup>Here we did not discuss some important observational effects, as we are just focusing on the comoving galaxy field. For example, we do not measure galaxy clustering distances directly in comoving space, rather we measure galaxy redshifts. Redshift is given by the Hubble expansion, as well as the peculiar velocities of galaxies. These lead to redshift space distortions: the observed redshift space overdensity of the galaxy field is given by the real space overdensity and a correction due to peculiar velocities [138]. In the distant-observer approximation, basically small angular separations between galaxies, on large-scales the correction depends on cosmological parameters via the logarithmic growth rate in linear theory  $f(z) = \frac{d \ln D}{d \ln a}$ , and on large-scales, the total observed galaxy clustering signal part only is  $P_{\text{rsd}} = P_m(k) [b_1 + f\mu^2]^2$ , where  $\mu$  is the angle between the line of sight and the  $\vec{k}$  vector. In particular,  $f(z) \approx \Omega^\gamma(z)$ , with  $\gamma \approx 0.55$  for  $\Lambda$ CDM, making this parameter critical for dark energy, as well as modified gravity models.

$$\delta_h(\vec{x}) = b_1 \delta_m(\vec{x}) + \frac{1}{2} b_2 \delta_m^2(\vec{x}) . \quad (2.164)$$

Then, at leading order, the halo power spectrum and bispectrum, given by gravitational effects only, in perturbation theory are

$$P_h(k) = b_1^2 P_m(k) , \quad (2.165)$$

$$B_h(k_1, k_2, k_3) = b_1^3 B_m(k_1, k_2, k_3) + b_1^2 b_2 (P_m(k_1) P_m(k_2) + 2 \text{ cyc}) , \quad (2.166)$$

where  $B_m$  is the gravitational non-linear matter bispectrum, taking the form of  $B_m(k_1, k_2, k_3) = 2P_l(k_1)P_l(k_2)F_2(\vec{k}_1, \vec{k}_2) + 2 \text{ cyc}$ .

Measuring only  $P_h$  does not allow us to distinguish between  $b_1$  and the amplitude of the matter fluctuations (as measured by  $\sigma_8$  for example). But as the bispectrum  $B_h$  can distinguish, thanks to  $B_m$ , between different triangle configurations in  $F_2$ , we can break the degeneracy.

A non-vanishing matter bispectrum beyond the non-linear gravitational one could signal the presence of non-Gaussian initial conditions [23]. This will be discussed further in Chapter 5.

Measuring the bispectrum is more complicated, and time-consuming than measuring a power spectrum, making it challenging to include it in standard cosmological analyses. Recently, deep learning methods (e.g. [103]) have been used to overcome these problems, although they still lack transparency in their interpretation. On the other hand, simple proxy estimators, as in [236], have been proposed to access non-Gaussian information from the large-scale structure. In Chapter 5 we will see another estimator that helps to constrain primordial non-Gaussianity.

Even without bispectrum measurements, we can still break degeneracies between cosmological parameters by combining multiple datasets in the same cosmological analysis. Cross-correlation science has been gaining importance in recent years. In Chapter ?? we will take a look at the example of a CMB lensing-galaxy cross-correlation. Due to its power, extracting CMB lensing from CMB data is an important task that we will explore in the next section.

## 2.8 Reconstructing modes with quadratic estimators

If we take a closer look at equations (2.131) for weak CMB lensing, and (2.147) for matter clustering, in the SPT picture, we will notice that they are quite similar in their form. They both have an observed field that has a first-order part, and higher-order mode couplings. We will see how we can use this picture to recover modes of the coupling field in an easy way. These can be then used for data analysis to extract cosmological parameters.

### Quadratic Reconstruction

We model the observed field  $X^i$  at a generic Fourier mode  $\vec{t}$  as given by an original statistically homogeneous and isotropic field  $Y$  coupled with an extra field  $Z$ , with a mode coupling  $\alpha_Z$ :

$$X(\vec{t}) = Y(\vec{t}) + \int_{\vec{t}'} \alpha_Z(\vec{t}', \vec{t}) Y(\vec{t}') Z(\vec{t} - \vec{t}') , \quad (2.167)$$

where  $\int_{\vec{t}'} \equiv \int \frac{d^N \vec{t}'}{(2\pi)^N}$ , for  $N$  dimensions.

If we now measure the correlation of  $X$  at two different  $\vec{t} \neq \vec{t}'$ , averaging over  $Y$ , and fixing  $Z$  (if we imagine  $Y$  and  $Z$  to be uncorrelated), and assuming  $Z$  is real ( $Z^*(\vec{t}' - \vec{t}) = Z(\vec{t} - \vec{t}')$ ),

$$\langle X(\vec{t}) X^*(\vec{t}') \rangle_Y = \left[ \alpha^*(\vec{t}, \vec{t}') C_{|\vec{t}|}^Y + \alpha(\vec{t}', \vec{t}) C_{|\vec{t}'|}^Y \right] Z(\vec{t} - \vec{t}') . \quad (2.168)$$

We see that the field has extra non-diagonal covariance, breaking statistical homogeneity.

We can write an estimator for the field  $Z$  that sums over pairs of modes with some weighting function  $g$

$$\hat{Z}(\vec{T}) = \int_{\vec{t}_1} \int_{\vec{t}_2} g_X^{YY}(\vec{t}_1, \vec{t}_2) X(\vec{t}_1) X(\vec{t}_2) \delta_D^N(\vec{T} - \vec{t}_1 - \vec{t}_2) . \quad (2.169)$$

We note that the weighting, in general, is *not* symmetric in its arguments.

We have some freedom on the weighting. Usually, the most important request for the estimator is to not be biased,

$$\langle \hat{Z}(\vec{T}) \rangle_Y = Z(\vec{T}) , \quad (2.170)$$

and this gives a constraint on  $g$ .

The covariance of the  $Z$  estimator can be split into two parts, one Gaussian and one non-Gaussian

$$\begin{aligned} \langle \hat{Z}(\vec{T}) \hat{Z}^*(\vec{T}') \rangle - \langle \hat{Z}(\vec{T}) \rangle \langle \hat{Z}^*(\vec{T}') \rangle &= (2\pi)^N \delta_D^N(\vec{T} - \vec{T}') \left( \text{Cov}_G \left[ \hat{Z}(\vec{T}), \hat{Z}^*(\vec{T}') \right] \right. \\ &\quad \left. + \text{Cov}_{\text{NG}} \left[ \hat{Z}(\vec{T}), \hat{Z}^*(\vec{T}') \right] \right). \end{aligned} \quad (2.171)$$

The Gaussian part will be given by disconnected two point functions, using Wick's theorem,

$$\begin{aligned} \text{Cov}_G \left[ \hat{Z}(\vec{T}), \hat{Z}^*(\vec{T}') \right] &= \int_{\vec{t}} \int_{\vec{t}'} g_X^{YY}(\vec{t}, \vec{T} - \vec{t}) g_X^{YY*}(\vec{t}', \vec{T}' - \vec{t}') \\ &\quad \times \left[ \langle X(\vec{t}) X(\vec{T} - \vec{t}) X(\vec{t}') X(\vec{T}' - \vec{t}') \rangle - \langle X(\vec{t}) X(\vec{T} - \vec{t}) \rangle \langle X(\vec{t}') X(\vec{T}' - \vec{t}') \rangle \right] = \\ &= \int_{\vec{t}} \int_{\vec{t}'} g_X^{YY}(\vec{t}, \vec{T} - \vec{t}) g_X^{YY*}(\vec{t}', \vec{T}' - \vec{t}') \times \\ &\quad \times \left[ P^{XX}(\vec{t}) P^{XX}(\vec{T} - \vec{t}) (2\pi)^N \delta_D^N(\vec{t} - \vec{t}') (2\pi)^N \delta_D^N(\vec{T} - \vec{T}') + \right. \\ &\quad \left. + P^{XX}(\vec{t}) P^{XX}(\vec{T} - \vec{t}) (2\pi)^N \delta_D^N(\vec{T} - \vec{t} - \vec{t}') (2\pi)^N \delta_D^N(\vec{T} - \vec{T}') \right] = \\ &= (2\pi)^N \delta_D^N(\vec{T} - \vec{T}') \int_{\vec{t}} g_X^{YY}(\vec{t}, \vec{T} - \vec{t}) \times \\ &\quad \times \left[ g_X^{YY*}(\vec{t}, \vec{T} - \vec{t}) P^{XX}(\vec{t}) P^{XX}(\vec{T} - \vec{t}) + g_X^{YY*}(\vec{T} - \vec{t}, \vec{t}) P^{XX}(\vec{t}) P^{XX}(\vec{T} - \vec{t}) \right], \end{aligned} \quad (2.172)$$

where we use the property of Dirac delta functions in  $\mathbf{R}^N$   $f(\vec{x}) \delta_D^N(\vec{x} - \vec{x}') = f(\vec{x}') \delta_D^N(\vec{x} - \vec{x}')$ , and we define the power spectrum  $P^{XY}$  between two fields  $X, Y$ .

With this notation we want to capture the relevant cases of  $N = 2$ , where  $P(\vec{t}) \equiv C(\vec{l})$ , and  $N = 3$ , where  $P(\vec{t}) \equiv P(\vec{k})$ . The Gaussian part of the covariance matrix will contain contributions coming from the total power spectrum, which in general will include cosmic variance error due to finite number of modes and noise.

### 2.8.1 CMB lensing reconstruction

We will now apply the ideas of the previous section to CMB lensing reconstruction. As discussed previously, CMB lensing is a direct probe of the projected matter distribution in the Universe, containing information about the clustering and its geometry, and making this application significant. How, then, do we reconstruct the CMB lensing potential from the observed CMB?

We begin by writing the CMB lensed field in the first order in the lensing potential

$$T(\vec{l}) = T^u(\vec{l}) + \int \frac{d^2\vec{l}'}{(2\pi)^2} T^u(\vec{l}') K^\phi(\vec{l}, \vec{l}') , \quad (2.173)$$

where we remember equation (2.132).

This expression bears resemblance to (2.167). The lensing potential introduces couplings between different CMB modes. If we imagine the late large-scale structure to be uncorrelated from the primordial CMB, and in the limit of a fixed structure, averaged over statistical realizations of the primordial CMB, we have that for two different lensed (beam-deconvolved) temperature modes  $\vec{l}, \vec{l}'$  [129]

$$\langle T(\vec{l})T(\vec{l}') \rangle_{\text{CMB}} = f_{TT}(\vec{l}, \vec{l}') \phi(\vec{L})|_{\vec{L}=\vec{l}+\vec{l}'}, \quad (2.174)$$

where the function  $f_{TT}$  is

$$f_{TT}(\vec{l}, \vec{l}') = C_l^u(\vec{l} + \vec{l}') \cdot \vec{l} + C_{l'}^u(\vec{l} + \vec{l}') \cdot \vec{l}', \quad (2.175)$$

and we have used (2.173), statistical homogeneity in Fourier space, and  $\phi^*(\vec{L}) = \phi(-\vec{L})$  for reality of  $\phi$ .

The fluctuations are Gaussian but they become statistically anisotropic, with anisotropy proportional to the lensing potential calculated at some  $\vec{L}$ . Inspired by this formula, a naive (sub-optimal) estimator for the lensing potential is thus

$$\hat{\phi}(\vec{L}) \sim \frac{T(\vec{l})T(\vec{L} - \vec{l})}{f_{TT}(\vec{l}, \vec{L} - \vec{l})} \quad (2.176)$$

This is an unbiased estimator for  $\phi$ ,  $\langle \hat{\phi} \rangle_{\text{CMB}} \sim \phi$ , although it is a much noisier estimator than taking all the CMB pair modes with a difference equal to the CMB lensing mode to be reconstructed.

A better estimator than this, which exploits more modes, is

$$\hat{\phi}(\vec{L}) = \int \frac{d^2\vec{l}}{(2\pi)^2} T(\vec{l})T(\vec{L} - \vec{l})g_{TT}(\vec{l}, \vec{L} - \vec{l}) . \quad (2.177)$$

From the results of the previous section, and using that  $f_{TT}$  is real and symmetric in its arguments, we find that

$$g_{TT}(\vec{l}, \vec{L} - \vec{l}) = N^{TT}(\vec{L})F_{TT}(\vec{l}, \vec{L} - \vec{l}) = N^{TT}(\vec{L}) \frac{f_{TT}(\vec{l}, \vec{L} - \vec{l})}{2C_{|\vec{l}|}C_{|\vec{L}-\vec{l}|}}, \quad (2.178)$$

and the normalization, to ensure that  $\langle \phi(\vec{L}) \rangle = \phi(\vec{L})$ , is

$$N(\vec{L}) = \left[ \int_{\vec{l}} f_{TT}(\vec{l}, \vec{L} - \vec{l})g(\vec{l}, \vec{L} - \vec{l}) \right]^{-1} . \quad (2.179)$$

In reality, the CMB lensing estimator uses input CMB maps that contain a CMB part, an experimental noise part (due to finite resolution of the experiment and possibly the atmosphere), and foregrounds. Repeating the same exercise as above, and requiring that the estimator recovers the true lensing field with the minimum noise possible, we find that

$$g_{TT}(\vec{l}, \vec{L} - \vec{l}) = N^{TT}(\vec{L}) F_{TT}(\vec{l}, \vec{L} - \vec{l}) = N^{TT}(\vec{L}) \frac{f_{TT}(\vec{l}, \vec{L} - \vec{l})}{2C_{|\vec{l}|, \text{exp}} C_{|\vec{L}-\vec{l}|, \text{exp}}}, \quad (2.180)$$

where, now, we specify the total beam-deconvolved experimental CMB power spectrum,  $C_{|\vec{l}|, \text{exp}}$ .

The CMB lensing power spectrum can be estimated from data as follows

$$\langle \hat{\phi}(\vec{L}) \hat{\phi}^*(\vec{L}') \rangle = N^2(\vec{L}) \int_{\vec{l}} \int_{\vec{l}'} F(\vec{l}, \vec{L}') F(\vec{l}', \vec{L}) \langle \tilde{T}_{\text{exp}}(\vec{l}) \tilde{T}_{\text{exp}}(\vec{L} - \vec{l}) \tilde{T}_{\text{exp}}(\vec{l}') \tilde{T}_{\text{exp}}(\vec{L}' - \vec{l}') \rangle. \quad (2.181)$$

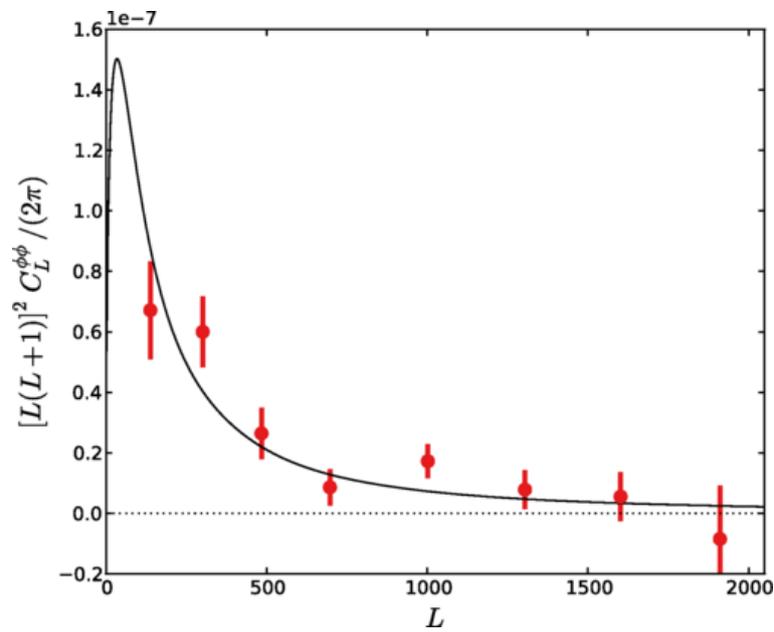
This CMB lensing power spectrum estimator is biased, with several unwanted contributions, such as the Gaussian part of the four-point function, present even in the absence of gravitational lensing. This can be seen by applying Wick's theorem to (2.181).

We can then write

$$\langle \hat{\phi}(\vec{L}) \hat{\phi}^*(\vec{L}') \rangle = (2\pi)^2 [C_L^{\phi\phi} + N_L + \dots] \delta_D^{(2)}(\vec{L} - \vec{L}'), \quad (2.182)$$

where  $\dots$  includes higher-order biases to the CMB lensing reconstruction (e.g. [256]).

In general, we will measure the same lensing potential from all the available CMB quadratic couplings, e.g.  $XY \in \{TT, TE, EE, EB\}$ . Formulae for the specific weights can be found in [130], although the unlensed power spectra there should be replaced by lensed power spectra, as suggested in [109], to cancel higher-order biases to the estimator that uses unlensed spectra. In Figure 2.12 we show an example of a measured CMB lensing power spectrum from the Atacama Cosmology Telescope CMB survey [256].



*Fig. 2.12* In red, the two-season ACTPol measured lensing power spectrum. In black, the theoretical best fit in  $\Lambda$ CDM CMB lensing power spectrum  $C_L^{\phi\phi}$ . Figure from [256].

# The Atacama Cosmology Telescope: A CMB lensing mass map over 2100 square degrees of sky and its cross-correlation with BOSS-CMASS galaxies

## Summary

*We construct cosmic microwave background lensing mass maps using data from the 2014 and 2015 seasons of observations with the Atacama Cosmology Telescope (ACT). These maps cover 2100 square degrees of sky and overlap with a wide variety of optical surveys. The maps are signal dominated on large scales and have fidelity such that their correlation with the cosmic infrared background is clearly visible by eye. We also create lensing maps with thermal Sunyaev-Zel'dovich contamination removed using a novel cleaning procedure that only slightly degrades the lensing signal-to-noise ratio. The cross-spectrum between the cleaned lensing map and the BOSS CMASS galaxy sample is detected at  $10\text{-}\sigma$  significance, with an amplitude of  $A = 1.02 \pm 0.10$  relative to the Planck best-fit  $\Lambda$ CDM cosmological model with fiducial linear galaxy bias. Our measurement lays the foundation for lensing cross-correlation science with current ACT data and beyond.*

## 3.1 Introduction

Along their paths to our telescopes, the photons of the cosmic microwave background (CMB) are deflected, or lensed, by the gravitational influence of the matter in our Universe. This leads to a remapping of the observed CMB anisotropies on the

sky described by  $T(\hat{n}) = T^u(\hat{n} + \vec{d})$ , where  $T$  and  $T^u$  are the lensed and unlensed temperature fields and  $\hat{n}$  is the line of sight. (Analogous expressions hold for the remapping of polarization  $Q$  and  $U$ .) The lensing deflection field  $\vec{d}(\hat{n})$  that describes the remapping depends on a weighted integral of the mass along the line of sight; although this integral extends to the last-scattering surface, most of the lensing signal arises between redshifts  $z = 0.5$  and  $z = 3$  [291, 153]. Since maps of the CMB lensing signal are sensitive to the total matter distribution, including dark matter, they contain a wealth of information about cosmology and fundamental physics [e.g., 152, 254, 212].

In this chapter, we present a CMB lensing map constructed from new observations from ACT, which will be useful for cross-correlation analyses.

Cross-correlation measurements can be used to break the degeneracy of galaxy bias (the factor relating the galaxy and matter density contrasts) and the amplitude of matter density fluctuations. This allows us to determine the amplitude of structure at different redshifts  $\sigma_8(z)$  [e.g., 96, 205, 100, 83] and hence probe physics such as dark energy, modified gravity, and neutrino mass. CMB lensing cross-correlations can also be used to constrain multiplicative biases in shear measurements [e.g., 276, 73, 108, 164, 231], measure cosmographic distance ratios [e.g., 127, 74, 186, 214], calibrate the masses of galaxy groups and clusters [e.g., 168, 35, 184, 211, 295, 220, 34, 220, 221], and probe astrophysics via the relation of dark to luminous matter [e.g., 255, 42, 279, 16, 208, 93, 107, 133, 199, 222]. However, a key challenge in such analyses is that CMB lensing maps reconstructed from temperature anisotropies can be contaminated by foreground emission and scattering [263, 118, 277, 75, 89], which can induce 10–20% level biases in the measured cross-correlation signal [200, 36]. For cross-correlations with low-redshift tracers, these foreground biases arise predominantly from the thermal Sunyaev-Zel’dovich (tSZ) residuals that lie in the map.

To solve this problem, in this paper we develop and implement a new cleaning method, building on [169] (hereafter MH18), in order to eliminate foregrounds from the tSZ effect in cross-correlations. The foreground removal in our method is achieved while preserving nearly all of the cross-correlation signal-to-noise.

We demonstrate the potential of our new foreground-cleaned CMB lensing maps, which overlap with a variety of optical surveys, by measuring a robust cross-correlation of these maps with Sloan Digital Sky Survey DR12 BOSS CMASS spectroscopic galaxies [223].

We also note that some analyses found a lower cross-correlation spectrum between CMB lensing and both low-redshift galaxies and weak lensing than expected from the *Planck* cosmology [e.g., 217, 163]. Testing this possible discrepancy with our new lensing maps provides further motivation for our analysis.

This chapter is structured as follows. Section 2 explains the theoretical background for our cross-correlation measurement. In Section 3 we present our data and discuss the new lensing maps constructed from ACT data. In Section 4 we discuss the construction of tSZ-free lensing maps. In Section 5 we present the cross-correlation measurement with CMASS BOSS galaxies, followed by a discussion of systematic errors in Section 6. The conclusions follow in the final section of the chapter. Two appendices explain the CMB map pre-processing and discuss, in more detail, the cleaning method used to remove the tSZ bias from the lensing maps.

## 3.2 Theoretical Background

The CMB lensing convergence field  $\kappa$ , which is related to the lensing deflection via  $\kappa = \frac{1}{2}\vec{\nabla} \cdot \vec{d}$ , is a direct measure of the projected matter field. In particular, the convergence can be shown to equal a weighted integral of the matter density perturbation along a line of sight with direction  $\hat{n}$ :

$$\kappa(\hat{n}) = \int_0^{z_*} dz W^\kappa(z) \delta(\chi(z)\hat{n}, z), \quad (3.1)$$

with  $z_*$  the redshift at the last-scattering surface,  $\delta$  the three-dimensional matter density contrast field at redshift  $z$ ,  $\chi(z)$  the comoving distance at redshift  $z$ , and the window response kernel  $W^\kappa$  for redshift  $z$  given by [e.g., 255]

$$W^\kappa(z) = \frac{3}{2H(z)} \Omega_{m,0} H_0^2 (1+z) \chi(z) \frac{\chi_* - \chi(z)}{\chi_*}, \quad (3.2)$$

where  $H(z)$  is the Hubble parameter as a function of redshift,  $H_0$  its value today,  $\chi_* = \chi(z_*)$ , and  $\Omega_{m,0}$  is the value of the matter density parameter today.

The 3D distribution of galaxies can provide an independent view of the matter distribution in combination with lensing, and one that can probe the time dependence of structure growth. (In contrast,  $\kappa$  is a projection of the matter field over a very wide range of redshifts and so cannot provide tomographic information.) The relevant cosmological field is the fractional number overdensity of galaxies in a direction  $\hat{n}$ , given by another weighted integral along the line of sight

$$\delta_g(\hat{n}) = \int_0^{z_*} dz W^g(z) \delta_g^{3D}(\chi(z)\hat{n}, z), \quad (3.3)$$

where  $\delta_g^{3D}$  is the three-dimensional galaxy distribution at redshift  $z$  and the window function  $W^g(z)$  is  $\frac{dn}{dz}(z)$ , the redshift distribution of galaxies in a galaxy survey,

normalized to unity.<sup>1</sup> In this work, we consider a spectroscopic galaxy survey with a redshift-binned sample such that the kernel  $W$  is only non-zero between  $z_i$  and  $z_f$ , with  $z_i, z_f$  the low and high redshifts defining the survey.

Since galaxies are biased tracers of the underlying matter distribution, the matter-galaxy power spectrum is

$$P_{mg}(k, z) = b_{\text{cross}}(k, z)P(k, z) , \quad (3.4)$$

where  $b_{\text{cross}}(k, z)$  is a general scale- and redshift-dependent clustering bias and  $P(k, z)$  is the matter power spectrum [41]. In our cross-correlation analysis, we explicitly choose the scales and redshift-range included such that the scale- and redshift-dependence of the galaxy bias is not large and  $b_{\text{cross}}(k, z) \approx b = \text{const}$ . We will consider multipoles  $L$  in the range  $100 < L < 1000$ ; this choice will be motivated in Section 5.

The cross-power spectrum of the two observables  $\kappa$  and  $g$  is directly related to the cosmological parameters of the underlying  $\Lambda$ CDM model. Using the Limber approximation [158], the expression for the cross-spectrum in the linear  $\Lambda$ CDM model is [e.g., 198]:

$$C_L^{\kappa g} = \int_0^{z_*} dz \frac{H(z)}{\chi^2(z)} W^\kappa(z) \frac{dn}{dz}(z) P_{mg} \left( k = \frac{L + \frac{1}{2}}{\chi(z)}, z \right) . \quad (3.5)$$

### 3.3 Lensing Maps from ACT Data Alone

We construct two CMB lensing maps. The first map, described in this section, uses ACT data alone. The second, described in the following section, also uses multi-frequency data from *Planck* in order to clean foregrounds.

#### 3.3.1 CMB maps for lensing analysis

The lensing convergence maps used in this work are constructed from CMB temperature and polarization data taken by the polarization-sensitive receiver on the Atacama Cosmology Telescope (ACT), a 6-meter CMB telescope operating in the Atacama desert in Chile [see e.g., 275, 62, 10]. The CMB field maps are obtained from observations made during seasons 2014–2015 in the 98 GHz and 150 GHz frequency bands; these maps will be made public, along with our lensing maps, in the upcoming ACT data release 4 (DR4). We will consider data coming from two regions of the

---

<sup>1</sup>We do not include magnification bias, since its magnitude is negligible given the low redshift range of the galaxy catalog used in this work.

sky, one referred to as BN (from the 2015 season, covering  $\approx 1633$  sq. deg. of the sky overlapping the SDSS BOSS northern field, with effective co-added white noise level of approximately  $\Delta_T = 21\mu\text{K-arcmin}$  for temperature and  $\Delta_P = \sqrt{2}\Delta_T$  for polarization), and the other referred to as D56 (seasons 2014–2015, covering  $\approx 456$  sq. deg. of the sky, with effective co-added white noise level of approximately  $\Delta_T = 10\mu\text{K-arcmin}$  for temperature and  $\Delta_P = \sqrt{2}\Delta_T$  for polarization).<sup>2</sup> Given the proximity of the maps to the equator and their moderate extent in declination, the flat-sky approximation is sufficient at our accuracy for constructing lensing maps; a simple estimate of the inaccuracy of this approximation gives no detectable effect for D56 and only a 1% multiplicative bias for BN. We do not use 2013 or 2016 observations in our analysis (even though the latter are part of DR4), because the 2013 observations cover too little sky area and the 2016 observations are still too shallow to contribute significant signal-to-noise to cross-correlation measurements.

We combine the per-season and per-frequency CMB maps presented in [62] to provide the input maps for our lensing estimator. The details of this procedure are described in Appendix (A.1), but we briefly summarize them here. We construct our CMB input maps by co-adding source-subtracted<sup>3</sup> maps from the two frequencies and two seasons of the data and convolving the result to a common beam after masking. In addition, we inpaint (fill with an appropriately correlated Gaussian random field) a 6-arcmin-radius circular area around bright compact sources and SZ clusters using the maximum likelihood method of [44]. This inpainting step serves to reduce foreground biases arising from bright sources and massive clusters. We note that the main difference from the map processing employed in [256] is that the different frequencies and seasons are coadded with weights that are local in Fourier space rather than real space; this is more optimal for multifrequency data due to the strong frequency dependence of the beams.

The results of our map construction and preparation process are masked, beam-deconvolved dimensionless CMB fluctuation maps of temperature  $T$  as well as  $Q$  and  $U$  polarization in each of the two sky regions. The  $Q$  and  $U$  polarization maps are transformed into  $E$ – $B$  polarization maps using the pure  $E$ – $B$  decomposition method outlined in [166]. As a final step in the preparation of the maps for lensing reconstruction, we follow the nominal analysis methodology of [62] to reduce the impact of ground contamination in the  $T$ ,  $E$  and  $B$  maps, filtering out all modes  $\vec{l} = (\ell_x, \ell_y)$  that have  $|\ell_x| < 90$  and  $|\ell_y| < 50$ . We also remove all modes that are outside the range of scales  $500 < \ell < 3000$  in order to restrict our lensing analysis to

<sup>2</sup>Atmospheric noise contributes a  $1/f$  component that is non-negligible and must be included when forecasting the signal-to-noise in the lensing map.

<sup>3</sup>See [170], [62], [10] for details.

scales where the ACT map-maker transfer function is small<sup>4</sup> and where contamination from foregrounds is small ( $\ell < 3000$ ).

As well as processing data, we also produce  $N = 511$  CMB simulations matching each of the CMB maps described above. These simulations are generated using the pipeline described in [62] and include primary CMB, lensing, noise and foregrounds. The foregrounds are Gaussian and spatially homogeneous and the noise is Gaussian but spatially inhomogeneous, as described in [62]. We use the simulations to test our lensing reconstructions, derive small transfer function corrections and construct covariance matrices, as described in the following sections of this chapter. To reconstruct lensing convergence maps from simulations we use the same pipeline that we apply to the data. We describe this lensing reconstruction pipeline in the following subsection.

### 3.3.2 Lensing reconstruction and validation

Exploiting the mode couplings induced by lensing, we reconstruct the lensing convergence field from our CMB maps with a minimum variance quadratic estimator [130]:

$$\bar{\kappa}^{XY}(\vec{L}) = A^{XY}(\vec{L}) \int \frac{d^2\vec{l}}{(2\pi)^2} X(\vec{l}) Y(\vec{L} - \vec{l}) f^{XY}(\vec{l}, \vec{L}), \quad (3.6)$$

where  $A^{XY}(\vec{L})$  is a normalization (derived from our fiducial cosmology) to ensure that the estimator is unbiased.  $f^{XY}(\vec{l}, \vec{L})$  is an optimal weighting function chosen to minimise the reconstruction noise of the estimator; it includes a Wiener filter for the CMB input fields  $X, Y$ . As in [256] we will consider only the pairs  $XY \in \{TT, TE, EE, EB\}$ , as the  $TB$  combination has negligible signal-to-noise. Expressions for the weighting function  $f$  and the theory normalization  $A$  can be found in [130], although following [109] we replace the unlensed spectra with lensed spectra in the weighting functions to cancel higher-order biases. A spurious signal on the largest scales of the reconstructed lensing map arises from non-lensing statistical anisotropy due to sky masks or inhomogeneous map noise; this spurious lensing “mean field” must be subtracted from Equation (3.6) [e.g., 195]. We calculate this mean field correction by generating 511 lensing reconstructions from simulations and averaging these reconstructions. We thus obtain the mean-field subtracted lensing convergence

<sup>4</sup>The map-maker transfer function is close to unity for  $\ell > 500$  in D56, but the deviation from unity may be as large as 10% in the BN analysis region between  $\ell$  of 500 and 600 [62, 10]. However, because of the fact that the lensing estimator only draws a small fraction of its statistical weight from multipoles  $500 < \ell < 600$  (less than 2%, see e.g. [239]), we expect an effect on lensing cross-correlations that is much smaller than the statistical uncertainty and is thus negligible.

estimator

$$\hat{\kappa}^{XY}(\vec{L}) = \bar{\kappa}^{XY}(\vec{L}) - \langle \kappa_s^{XY}(\vec{L}) \rangle_s, \quad (3.7)$$

where  $\kappa_s^{XY}(\vec{L})$  is the lensing reconstruction  $\bar{\kappa}^{XY}$  for the simulation realization  $s$  and the angle average  $\langle \rangle_s$  is over simulations.

We complete the lensing map by creating a minimum variance combination of the different types of quadratic estimators  $XY \in \{TT, TE, EE, EB\}$ ,

$$\hat{\kappa}_{\vec{L}}^{\text{MV}} = \sum_{XY} w^{XY}(\vec{L}) \hat{\kappa}^{XY}(\vec{L}), \quad (3.8)$$

where  $w^{XY}(\vec{L})$  are minimum-variance weights.

Finally, the particular form of the normalization  $A^{XY}(\vec{L})$  used in Equation (3.6) is valid for CMB maps with periodic boundaries. This is clearly an idealization; for example, using masked CMB maps introduces spurious gradients at the mask boundary [116], changing the form of the correct lensing normalization (although this effect is reduced by apodization). We capture this and other non-idealities by introducing an extra multiplicative normalization function  $r^{\text{MC}}(L)$ .

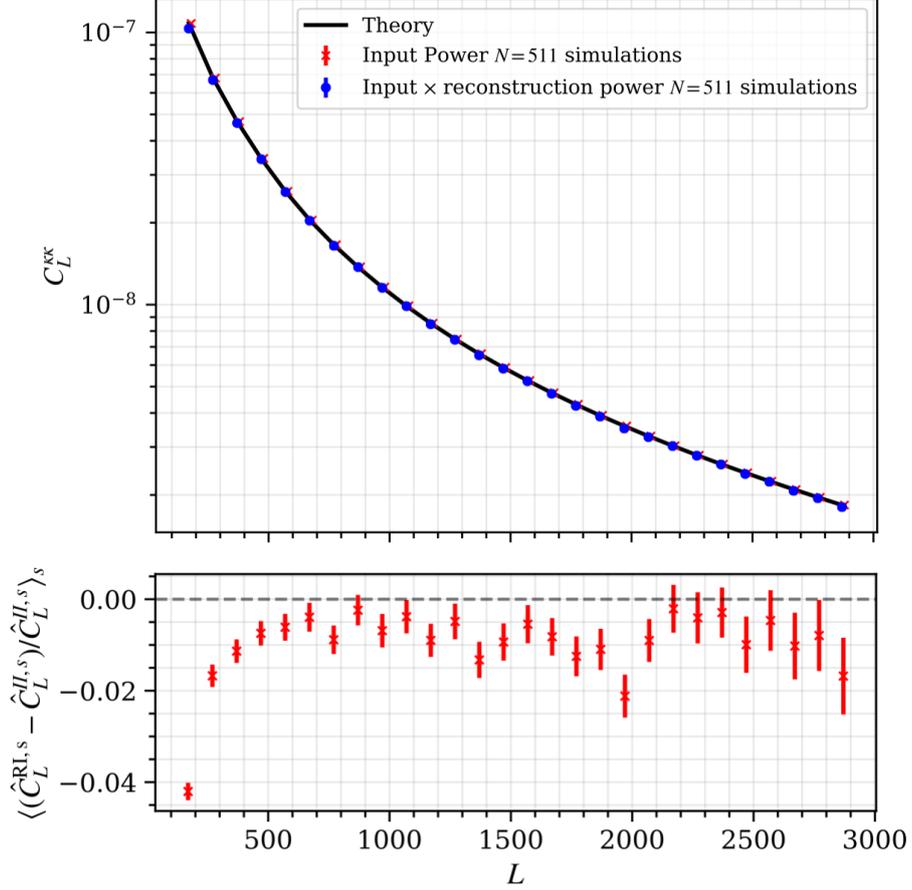
To calculate this function, we cross-correlate our  $N = 511$  reconstructed lensing simulations  $\hat{\kappa}_s^{\text{MV}}$  with the true input lensing convergence field  $\kappa_s$  used to generate the simulations,<sup>5</sup> obtaining the reconstruction-input cross-spectrum  $\hat{C}_L^{\text{RI},s}$ . We compare this cross-spectrum with the auto-spectrum of the input convergence field  $\hat{C}_L^{\text{II},s}$ . Taking the ratio of averages over the  $N$  sims  $\langle \hat{C}_L^{\text{II},s} \rangle_s / \langle \hat{C}_L^{\text{RI},s} \rangle_s$ , we obtain a one dimensional binned function of  $L = |\vec{L}|$ , where  $L_{\text{min}} = 20$ ,  $L_{\text{max}} = 3000$ , and  $\Delta L = 100$ . We then interpolate this over a two dimensional grid to get the final isotropic correction function  $r^{\text{MC}}(\vec{L})$  that we apply to the lensing maps to obtain the MC corrected minimum variance lensing maps

$$\hat{\kappa}_{\vec{L}} = r^{\text{MC}}(\vec{L}) \hat{\kappa}_{\vec{L}}^{\text{MV}}. \quad (3.9)$$

If our pipeline is estimating the lensing signal reliably, the Monte-Carlo based normalization correction of Equation (3.9) should only require a rescaling of order a few percent. To validate our pipeline, we therefore test whether our lensing map is nearly correctly reconstructed even in the absence of Monte-Carlo renormalization.

In Figure (3.1) we show a comparison between  $\langle C_{LL}^{\text{RI},s} \rangle_s$  and  $\langle C_L^{\text{II},s} \rangle_s$  for the D56 patch without the Monte-Carlo normalization (this figure uses foreground-cleaned ACT+*Planck* lensing maps that we will introduce in the next section, but the residuals

<sup>5</sup>To mimic the processing of the reconstructions we mask  $\kappa_s$  with the square of the data-mask, as this enters twice in the quadratic lensing estimator used to reconstruct the lensing simulation.

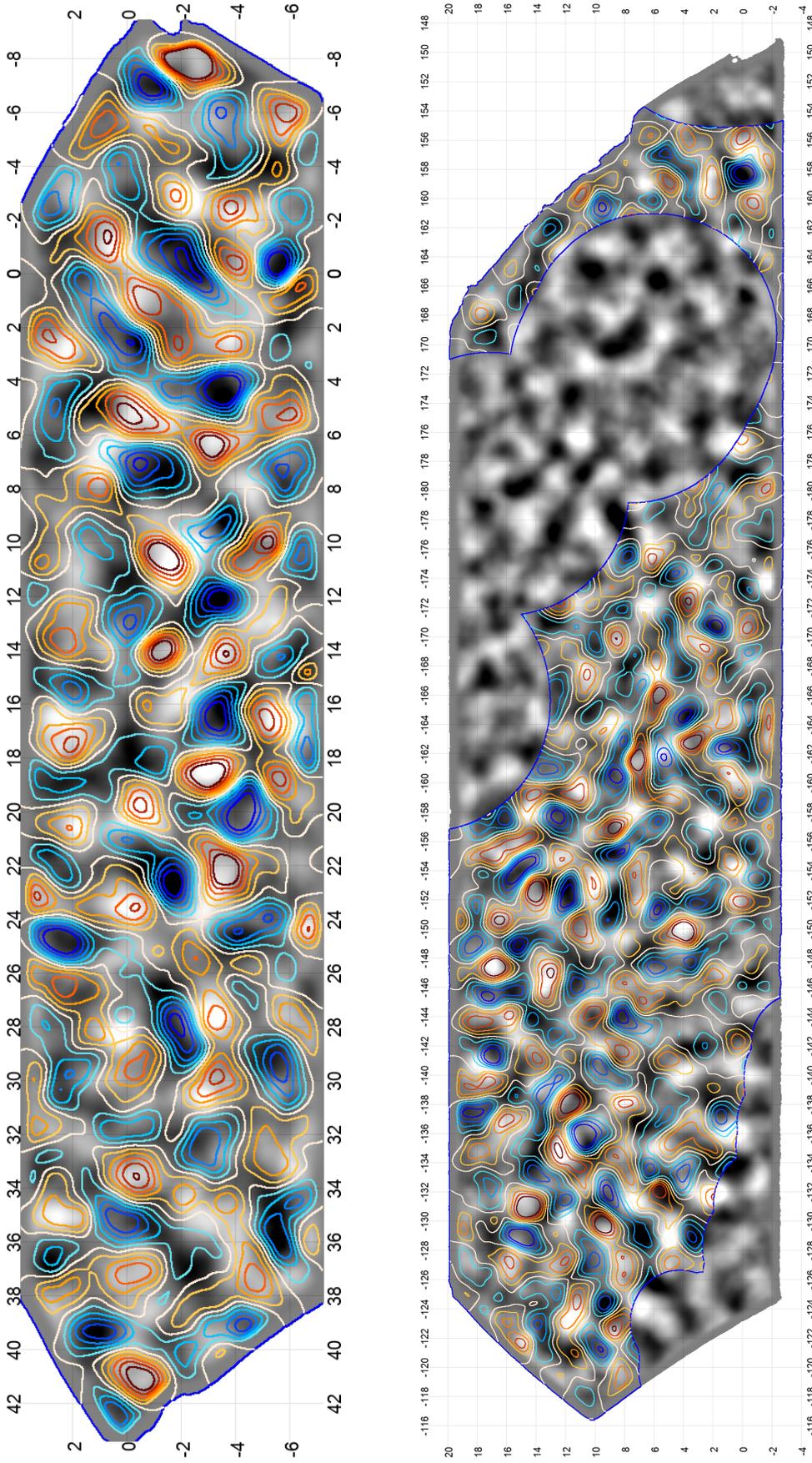


*Fig. 3.1* Verification of our lensing reconstruction pipeline for the tSZ-free lensing maps (shown for the D56 patch). We plot the average cross-spectrum of the reconstructed lensing maps with the input lensing simulations (blue dots), the average power spectrum of the input lensing simulations (red crosses) and a binned lensing power theory curve in black. (The BN patch gives quantitatively similar results.) The bottom panel shows the fractional difference of the input-reconstruction cross-correlation relative to the input lensing power. The ACT only simulations give residuals of similar magnitude. From the good agreement of the input-reconstruction cross-correlation with the input lensing power, we can see that the true lensing signal in the simulations is recovered within percent-level accuracy; we absorb only a small correction into a simulation-based re-normalization.

for the ACT-only maps are similar). We recover the signal with only percent-level deviations (which implies that  $r^{\text{MC}}(\vec{L})$  is within a few percent of unity); this gives confidence that our pipeline is functioning correctly. We obtain quantitatively similar results for the BN patch.

### 3.3.3 Visualization of the maps and their correlation with large-scale structure

An image of the ACTPol CMB lensing maps is shown in Figure (3.2). The maps have been Wiener filtered to show the signal-dominated scales (roughly 1 degree or larger for BN and 0.5 degrees or larger for D56) and have been converted to maps of the lensing potential using the appropriate filtering (given by switching between the lensing convergence  $\kappa$  and the lensing potential  $\phi$  with  $\kappa = L^2\phi/2$ ). We also overplot contours of Cosmic Infrared Background (CIB) emission obtained from the GNILC *Planck* component separated maps [213]; the CIB maps have the same filtering applied as the lensing ones. In the BN region, we mask the CIB map using the *Planck* PR2 Commander high-resolution map of thermal dust emission [210]. The mask is made by thresholding the dust map such that it covers regions of the CIB map that have visibly low power due to dust contamination; we only use this mask for the visualization of Figure (3.2). The CIB arises from similar redshifts as CMB lensing and hence is known to be highly correlated with lensing [265, 120, 209]. Indeed, even by eye a high correlation of our lensing maps and the CIB is visible. This illustrates the fact that our lensing maps are signal-dominated over a range of large scales and are a faithful tracer of the mass distribution (for other highly signal-dominated CMB lensing maps, see also [289]).



*Fig. 3.2* Map of the reconstructed lensing potential in the D56 region (upper panel) and the BN region (lower panel) after Wiener filtering, shown in greyscale. (The lensing maps shown are the tSZ-cleaned maps combining *Planck* and ACT, although the ACT-only lensing maps appear similar.) Overlaid, we also show contours of an identically filtered but completely independent cosmic infrared background map (*Planck*: GNILC 545 GHz). Since the correlation between CMB lensing and the cosmic infrared background (CIB) is very high and since our CMB lensing map has high signal-to-noise ratio on large scales, the correspondence between the lensing potential and the CIB can be seen clearly. Parts of the CIB map contaminated by Galactic dust have been masked in the BN CIB contours for this visualization, using a mask derived from the *Planck* PR2 Commander thermal dust emission map.

### 3.4 Foreground-mitigated lensing maps with new cleaning methods

CMB temperature maps contain secondary anisotropies not only from lensing, but also from tSZ, CIB (Cosmic Infrared Background), kSZ (kinetic Sunyaev-Zel'dovich), and other foreground contributions arising from a wide range of redshifts. The lensing estimator is sensitive to these extragalactic foregrounds [see 277, 202, 89], which can be problematic: foreground contamination which has leaked through the lensing estimator can correlate with the galaxy distribution, giving spurious biases to cross-correlation measurements. It is important to mitigate these foregrounds in temperature, as many current- and next-generation lensing maps will still depend to a large extent on temperature data, rather than on polarization. Indeed, for our current dataset, the temperature ( $TT$ ) lensing estimator still provides the dominant contribution ( $> 50\%$ ) to our minimum-variance lensing estimate of Equation (3.8).

One of the primary goals of making a lensing map is to enable cross-correlation science. For low- $z$  large-scale-structure tracers, such as the CMASS galaxies used in later sections of this chapter, the main contribution to the cross-correlation bias comes from the tSZ contamination of the temperature maps [277, 36, 169]. The tSZ is most important because, while the tSZ and the CIB can both be significant contaminants, the CIB only weakly correlates with low- $z$  galaxies (as only a small fraction of the CIB arises from low redshifts).

The observed, SZ-contaminated temperature map, denoted  $T_{\text{with-sz}}$ , now includes an SZ contribution  $T_{\text{tSZ}}$ , so that  $T_{\text{with-sz}} = T_{\text{cmb}} + T_{\text{tSZ}}$ .<sup>6</sup> When inserting this CMB map into a quadratic lensing estimator  $\hat{\kappa}(T_{\text{with-SZ}}, T_{\text{with-SZ}})$  and cross-correlating the resulting lensing map with a galaxy map  $g$ , the cross-correlation is now biased by a new bispectrum term of the form  $\langle gT_{\text{tSZ}}T_{\text{tSZ}} \rangle$ .

For typical cross-correlations, this effect can be significant, giving biases up to a 10–20% level on large scales [200, 36]. The shape of the bias on large scales is typically similar to that of the signal itself; the sign of the effect is generically negative on large and intermediate scales  $L < 1000$  (with a positive bias only arising on very small scales), so that a cross-correlation with a tSZ-contaminated lensing map is biased low.<sup>7</sup>

<sup>6</sup>The observed temperature map clearly also has other contributions in addition to  $T_{\text{cmb}}$  and  $T_{\text{tSZ}}$ , but our focus here will be just on these two components.

<sup>7</sup>A physical explanation for this negative bias effect is the following. Consider a direction in which there is a long-wavelength overdensity. Due to non-linear evolution and mode coupling, small-scale tSZ fluctuations are also enhanced in this direction, which increases the CMB temperature power at small scales,  $l > 2000$ . This excess small-scale power is similar in effect to an overall ‘shift’ of the

Since low cross-correlations were found in several analyses [e.g., 217]; it is interesting to consider if this type of contamination could have an impact on previously published cross-correlation measurements. However, we note that most of the analyses with low cross-correlations used *Planck* lensing maps. For *Planck*, such foreground biases are expected to be much less problematic (due to the lower experimental angular resolution).

### 3.4.1 A new tSZ-free estimator

To account for the potential problem of tSZ contamination, we attempted to use the method of MH18 to remove foreground contamination. However, this method did not perform as well as expected. We therefore developed a new foreground-cleaned lensing estimator, extending and revising the MH18 method; we will explain the relevant details in the following paragraphs.

The basic goal of our foreground-cleaning approach is to remove foreground contamination without assuming a model for the foregrounds' statistical properties, relying instead on the fact that the foregrounds' frequency dependence differs from that of the CMB. A simplistic frequency cleaning of the CMB maps, however, typically degrades the lensing signal-to-noise. MH18 uses the standard lensing convergence quadratic estimator written in real space in a form where a gradient and a non-gradient field can be distinguished [e.g., 124, 153]. Usually, for the temperature quadratic estimator  $\hat{\kappa}(T_1, T_2)$ , the two fields  $T_1, T_2$  are chosen to be identical. However, one may, of course, use two different CMB temperature maps in the estimator; the two maps could be processed differently or even come from different surveys. In particular, since the spectral energy distribution (SED) of the tSZ effect is known to high accuracy (barring relativistic and multiple-scattering effects), CMB maps made from multi-frequency data that explicitly null or deproject the tSZ can be made. Such maps generally have higher noise. In the procedure suggested by MH18, it is pointed out that even if only one of the two fields in the quadratic estimator is free from tSZ, then the resulting lensing map cross-correlation will still have zero tSZ contamination, while the noise increase due to foreground cleaning will only be moderate (since only one noisy cleaned map is used, instead of two). One way of understanding this is to note that, since the cross-correlation bias arises from a foreground-foreground-galaxy bispectrum  $\langle gT_{\text{tSZ}}T_{\text{tSZ}} \rangle$ , nulling even one of the foreground fields sets the whole bispectrum  $\langle gT_{\text{tSZ}}0 \rangle$  to zero, which gives an effectively bias-free cross-correlation measurement.

---

primary CMB towards smaller scales. The lensing estimator interprets this locally as arising from demagnification due to a matter underdensity: cross-correlating this spurious underdensity lensing signal with the distribution of galaxies (which trace the overdensity) therefore results in a negative cross-correlation[277].

We denote this foreground-cleaned MH18 estimator as  $\hat{\kappa}(T_{\text{no-tSZ}}, T_{\text{with-tSZ}})$  (where the first map is the gradient field in the lensing estimator). Despite the use of a noisier tSZ-deprojected map in one field of the quadratic estimator, the loss in signal-to-noise in constructing this foreground-free lensing map was claimed in MH18 to be only  $\approx 5\%$ .

However, when implementing the MH18 estimator, we found that the actual lensing map noise obtained in both simulations and in data was larger for  $L < 800$  (by more than an order of magnitude at  $L \approx 100$ , see Figure (A.1)) than the noise forecast presented in MH18. The explanation for this result is the following: in MH18 a simplified formula for the noise forecast was used (namely assuming the noise is equal to the normalization, i.e.  $N_L \propto L^2 A_L$ ); however, this is only valid if the weights in the estimator are minimum-variance. As detailed in Appendix (A.2), the MH18 estimator does not use minimum-variance weights, which explains why the true noise we find is larger than the simplified forecast results. We note that the MH18 forecast is however accurate for cluster scales, where the gradient approximation holds in the squeezed limit [124, 221].

To solve the problem of increased noise on large scales, we propose a new ‘symmetrized’ cleaned estimator, in which we coadd the  $\hat{\kappa}(T_{\text{no-tSZ}}, T_{\text{with-tSZ}})$  MH18 estimator with a version where the two fields have been permuted,  $\hat{\kappa}(T_{\text{with-tSZ}}, T_{\text{no-tSZ}})$ . In particular, we define  $\hat{\kappa}_{\text{symm, tSZ-free}}^{TT} = \sum w_\alpha(\vec{L}) \hat{\kappa}_\alpha(\vec{L})$  with weights

$$w_\alpha(\vec{L}) = \frac{\sum_\beta N_{\alpha\beta}^{-1}(\vec{L})}{\sum_{\gamma,\beta} N_{\gamma\beta}^{-1}(\vec{L})}, \quad (3.10)$$

where  $\alpha \in \{(T_{\text{no-tSZ}}, T_{\text{with-tSZ}}), (T_{\text{with-tSZ}}, T_{\text{no-tSZ}})\}$  and  $N^{-1}$  is the inverse  $2 \times 2$  covariance matrix taking into account the cross-correlation between the two estimators.

The resulting  $\hat{\kappa}_{\text{symm, tSZ-free}}^{TT}$  map retains the property that the resulting cross-correlation with large-scale structure is unbiased, but the lensing map now has significantly lower noise: in fact, we find that our method appears to effectively recover the original forecast results of MH18, primarily due to the cancellation of anti-correlated noise on large scales from each of the two terms in the new estimator. Details can be found in Appendix B.

### 3.4.2 Application to data

The above technique requires maps of the CMB in which the tSZ signal has been deprojected (i.e., nulled) using multi-frequency data. Such maps were presented in

[170]; these maps were constructed by combining *Planck* and ACT<sup>8</sup> data using an internal linear combination (ILC) algorithm. We use the constrained ILC CMB map (with tSZ deprojection) and the standard ILC CMB map (with no deprojection)<sup>9</sup> from that analysis as the two input maps for the symmetrized cleaned lensing estimator  $\hat{\kappa}_{\text{symm, tSZ-free}}^{TT}$  described above; we thus create new foreground-cleaned temperature lensing maps.<sup>1011</sup>

The maximum CMB multipole,  $\ell_{\text{CMB}}^{\text{max}} = 3000$ , typically used in CMB lensing analysis is motivated by the desire to reduce contamination from foregrounds such as the tSZ. Since the tSZ bias is nulled in this new estimator, it is plausible that this maximum multipole is unnecessarily conservative and can be increased, thus improving the signal-to-noise of the estimator. Motivated by this possibility, we increase our maximum multipole for the tSZ-free TT estimator map somewhat, to  $\ell_{\text{CMB}}^{\text{max}} = 3350$ ; we perform a null test (see next section) to test for problematic contamination from other foregrounds such as CIB or kSZ. (This type of contamination becomes large when we use a higher  $l_{\text{max}}$ , such as 3500 and 4000, causing null-test failures; for this reason, we choose to only modestly increase  $\ell_{\text{CMB}}^{\text{max}}$  to 3350.) Furthermore, since the ILC maps include information from *Planck* for  $\ell < 500$ , we also relax the minimum multipole cut from  $\ell_{\text{CMB}}^{\text{min}} = 500$  to  $\ell_{\text{CMB}}^{\text{min}} = 100$ , providing additional gains in signal-to-noise.

We then create a foreground-cleaned minimum-variance lensing map as in Equation (3.8). The coadding procedure is the same as for the ACT-only lensing map, except that temperature lensing is now obtained from the tSZ-free symmetric estimator  $\hat{\kappa}_{\text{symm, tSZ-free}}^{TT}$ . We successfully repeat the lensing validation described in Section 3.3 with our new foreground cleaned estimator; the results are shown in Figure (3.1).

<sup>8</sup>Despite including *Planck* data, in these maps, the small-scales relevant for lensing are dominated by the ACT 148 GHz and 97 GHz channels.

<sup>9</sup>We use version v1.1.1 of the maps for which bandpass corrections for the tSZ response may not be accurate at the few percent level at the map-level. However, since the tSZ bias is at most 20% in power, tSZ-cleaned cross-correlations are only affected at the 1% level, an order of magnitude below the statistical sensitivity of this work.

<sup>10</sup>Before applying the lensing estimator to these ILC maps we also inpaint SZ clusters as described for the ACT-only maps.

<sup>11</sup>We note that on CMB small scales ( $l_{\text{CMB}} \sim 3000$ ), our multi-frequency cleaning for tSZ deprojection is primarily achieved through the combination of 90 and 150 GHz channels from ACT, as the *Planck* data lacks useful information for  $l_{\text{CMB}} > 2200$ . In the tSZ deprojected leg, the CMB noise is very high on these small scales, as there are just two useful frequencies, and so we effectively do not use them for lensing reconstruction. Indeed, the gradient cleaning estimator picks most of the information from  $\langle T_{\text{low}} T_{\text{high}} \rangle$ , where  $l_{\text{CMB,low}} \leq 2200$  and the  $l_{\text{CMB,high}}$  can be up to some  $l_{\text{CMB,max}}$ , e.g. of 3000.

## 3.5 Galaxy cross-correlation measurement

In the previous sections, we have introduced two types of CMB lensing maps, which will be publicly available as part of the upcoming data release DR4 associated with [10] and [62]. As an example of their utility, we cross-correlate these lensing maps with galaxies from the BOSS survey’s CMASS galaxy catalog.

### 3.5.1 The CMASS galaxy map

We use the CMASS galaxy catalog (with redshifts  $z \in [0.43, 0.7]$ ) provided by the DR12 release of the BOSS spectroscopic survey<sup>12</sup> to construct a galaxy overdensity map. Given a pixel  $\vec{x}$ , we estimate the galaxy overdensity as

$$\delta_g(\vec{x}) = \frac{\sum_{i \in \text{unmasked } \vec{x}} w_i}{\frac{1}{N} \sum_{i, \text{all unmasked}} w_i} - 1, \quad (3.11)$$

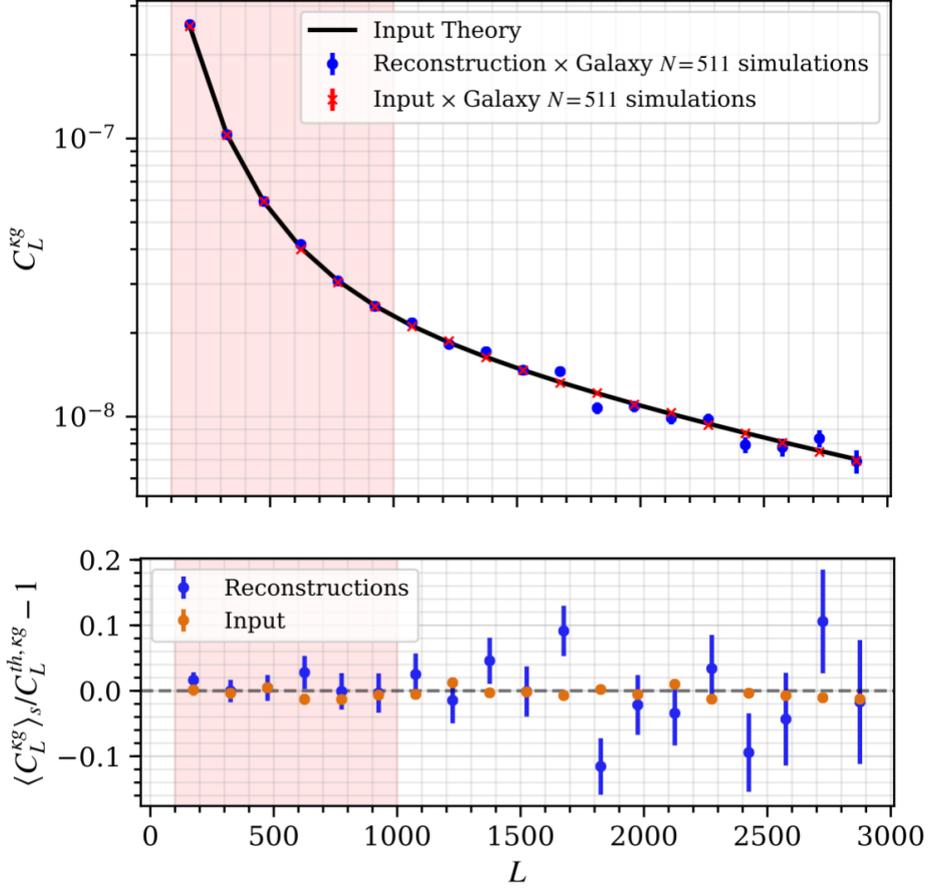
where  $N$  is the number of unmasked pixels (see below) and following [217, 186] each galaxy  $i$  inside the pixel  $\vec{x}$  is weighted according to

$$w = (w_{\text{noz}} + w_{\text{cp}} - 1)w_{\text{see}}w_{\text{star}}, \quad (3.12)$$

where  $w_{\text{noz}}$  accounts for redshift failures,  $w_{\text{cp}}$  for fiber collisions,  $w_{\text{star}}$  for bright star contamination and  $w_{\text{see}}$  for effects of seeing.

The galaxy mask used to mask pixels is created using ‘random catalogs’ provided by the BOSS collaboration; these catalogs contain a dense sampling of sky locations proportional to the survey conditions but not to any cosmological galaxy clustering signal. The random catalogs are mapped to a number density count map (created setting  $w = 1$ ) and then smoothed with a Gaussian beam with a width corresponding to a standard deviation of 2 arcminutes. To obtain the final mask, we then set to zero the regions of the smoothed randoms’ counts below a threshold of  $10^{-3}$ . The above choices are made so as to preserve survey information without picking up fluctuations in the random sampling. Our baseline analysis accounts for the effect of this mask simply by applying an overall scaling factor which compensates for the loss in power due to zeroed regions, as described in the next section. In general, the mask can also cause coupling of Fourier modes of the map leading to a modification of the estimated power spectrum. Although these effects are expected to be small since our mask is smooth, we test the impact of the mask on our cross-correlation measurement.

<sup>12</sup><http://www.sdss3.org/surveys/boss.php>



*Fig. 3.3* Lensing reconstruction test, as for Figure (3.1), but now correlating with a simulated galaxy field instead of the input lensing convergence field. The shaded region shows the multipole range used for the cosmological analysis. The lower panel shows the fractional difference with respect to the theory curve. (The BN patch gives similar results.)

We validate the treatment of the galaxy mask by applying it to mock Gaussian galaxy overdensity simulations which are correlated with the lensing signal according to a theoretical cross-spectrum with a fiducial bias  $b = 2$ . We verify that the cross-power spectrum measured from these simulations, with a multiplicative correction for the mask as described in the next section, reproduces the original input theory cross-correlation signal. As shown in Figure (3.3) we recover  $C_l^{ksg}$  to better than five percent over the cosmological analysis range, with no indication of an overall bias.

### 3.5.2 Extracting power spectra and obtaining the covariance matrix

Having constructed CMB lensing and galaxy maps we measure their cross-power spectra. Binned cross-power spectrum measurements are obtained using the following estimator valid for statistically isotropic fields,

$$\hat{C}_{L_b}^{\kappa g} = \frac{1}{w^{\kappa g}} \frac{1}{N_{\mathcal{A}}} \sum_{\vec{L} \in \mathcal{A}} \kappa_{\vec{L}}^{obs} g_{\vec{L}}^{obs*} \quad (3.13)$$

where  $\mathcal{A}$  is an annulus in the Fourier plane with average radius  $L = |\vec{L}|$ ,  $N_{\mathcal{A}}$  gives the number of modes in this annulus, and  $w^{\kappa g}$  is a correction factor due to masking that depends on the masked fields taken in consideration. For a slowly varying window function this is given by

$$w^{\kappa g} = \langle W_{\kappa}^2(\vec{x}) W_g(\vec{x}) \rangle, \quad (3.14)$$

where  $W_{\kappa}$  is the mask we apply to our CMB map before lensing reconstruction,  $W_g$  is the mask applied to our galaxy overdensity map, and the average is performed over pixels. Two powers of the CMB mask appear in the correction above because the lensing reconstruction is a quadratic estimator involving two powers of the CMB map.<sup>13</sup>

We obtain the covariance matrix for the cross-spectra from simulations as follows:

$$\hat{C}_{L_b, L'_b} = \langle (C_{L_b}^S - \langle \hat{C}_{L_b}^S \rangle_S) (C_{L'_b}^S - \langle \hat{C}_{L'_b}^S \rangle_S)^T \rangle_S \quad (3.15)$$

where the column power spectrum vector is  $\hat{C}_{L_b}^S = (C_{L_b}^{\kappa g S})^T$  and the average is over the simulations  $S$ .

To calculate this matrix, we cross correlate the  $N = 511$  lensing reconstruction simulations with the QPM mock catalogs of CMASS galaxies [285].<sup>14</sup> The cosmological signals in these simulations and catalogs are uncorrelated. We expect this not to be problematic because the uncorrelated part of the cross-correlation error dominates over the sample variance contribution. We verify this by calculating Gaussian theory standard errors with and without the  $(C_{L_b}^{\kappa g})^2$  sample variance term that arises from the presence of correlated structures, finding sub-percent level agreement between the two calculations.

<sup>13</sup>To avoid confirmation bias we did not plot a  $y$ -axis scale or overplot a theory curve over our cross-spectrum measurement until all the null tests and systematics checks, described in Section 3.6, had been successfully passed.

<sup>14</sup>Although more realistic mocks are available, the QPM mocks are sufficiently accurate for our purposes, i.e. to calculate error bars and verify our cross-correlation signal at the 10% level.

The inverse covariance matrix obtained from  $N$  simulations is calculated as in [111]:

$$\widehat{C}^{-1} = \beta \hat{C}^{-1}, \quad (3.16)$$

where  $\beta = \frac{N-p-2}{N-1}$  with  $p$  the number of angular bins.

Finally, we note that some care is required when choosing the range of scales  $L_{\min} < L < L_{\max}$  which we use in our analysis. Our theoretical model is expected to break down on smaller scales, since we are assuming a simple scale-independent linear galaxy bias, ignoring baryonic feedback on the matter power spectrum and also assuming that the non-linear matter power spectrum derived from HMCode [182], implemented in CAMB, is reliable. We therefore initially pick a range of scales based on the cross-correlation measurement; we set the requirement that the difference between a cross spectrum obtained from linear theory and one obtained from a non-linear power spectrum (HMCode), assuming a linear constant bias, should not be larger than the  $1\text{-}\sigma$  uncertainty for our cross-spectrum measurement. In this way, we obtain that the appropriate cutoff is approximately  $L_{\max, \kappa g} = 1000$ .<sup>15</sup>

In addition to the small-scale cuts described above, we also wish to avoid systematic errors which enter on large, degree-angular scales. On the galaxy side, such systematic errors include depth and selection function variations over the survey footprint; on the CMB lensing side, the main large-scale limitation is the challenge in simulating and subtracting the mean-field term sufficiently accurately, since it grows rapidly towards very low  $L$  ( $L < 50$ ). While many systematics are nulled in cross-correlation, they could induce additional variance, and to be safe we choose  $L_{\min} = 100$  for our analysis; at this scale, the power spectrum of the mean field is still smaller than that of the signal.

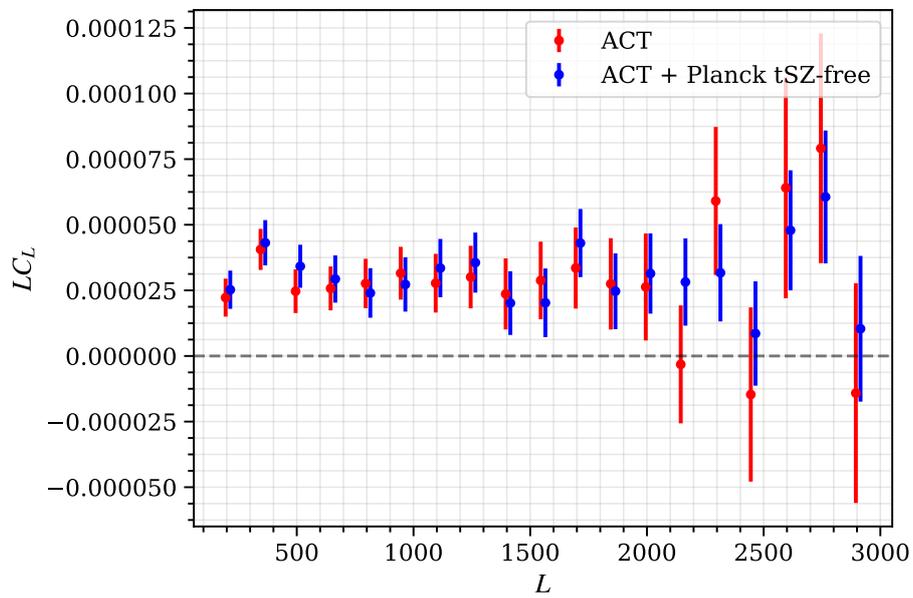
For our measurements, we choose a binning of  $\Delta L = 150$ ; with this binning, we find that the correlations between different bandpowers are not strong ( $< 13\%$ ).

### 3.5.3 Galaxy cross-correlation: results

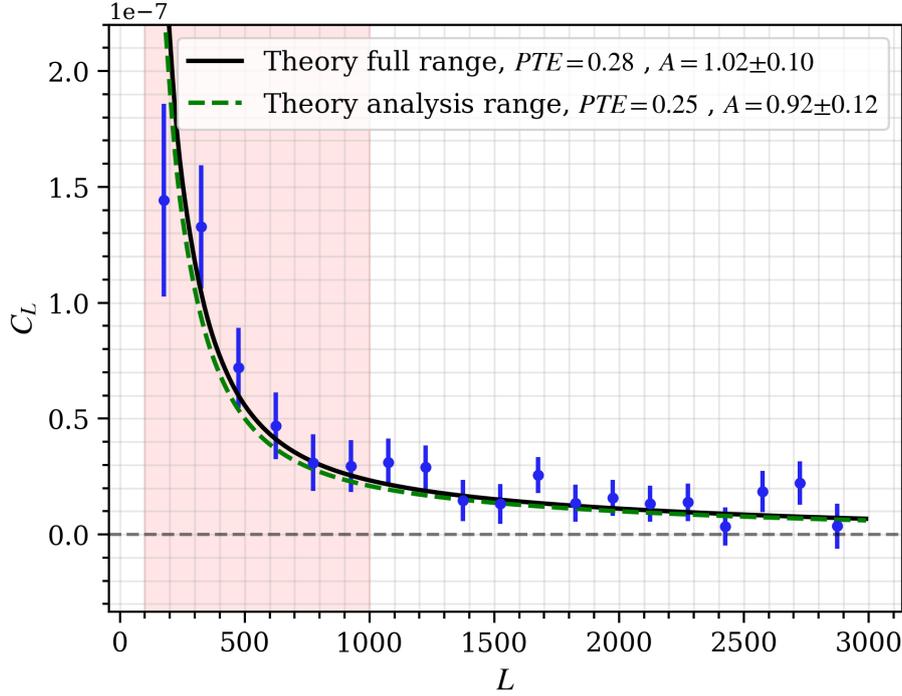
In Figure (3.4) we show the new tSZ-free CMB lensing–galaxy cross-correlation measurement. We also show the same cross-correlation with the ACT-only lensing maps, which have not been cleaned of tSZ.

A small shift between the bandpowers can be seen. It appears to match the form expected from bias due to tSZ in the ACT-only maps, *i.e.*, a deficit on large scales

<sup>15</sup>Even if this is beyond the non linear scale given at redshift  $z \sim 0.57$ , the mid-redshift of the CMASS catalog, the relatively large  $1\sigma$  uncertainty in the cross correlation at  $L \sim 1000$  implies that we are insensitive to the difference between linear and non-linear theory at that scale. Furthermore, as shown in Figure (3.5), including wave numbers beyond  $L \sim 1000$  only very slightly affects a simple cosmological analysis.



*Fig. 3.4* The cross-correlation between CMASS galaxies and CMB lensing convergence reconstructions from ACT. The cross-correlation measurements in the D56 and BN patches are coadded to obtain these results. The red points show results using an ACT-only lensing map, the blue points show results using a lensing map that has been tSZ cleaned. The multipole values for different versions of lensing maps are slightly offset for visualisation purposes. See Figure (3.5) for comparison with a theory curve fit to the cleaned measurement.



*Fig. 3.5* This plot shows our main CMB lensing – BOSS galaxy cross-correlation measurement with the ACT+*Planck* tSZ free lensing maps (blue points). A *Planck*-cosmology (and fiducial galaxy bias  $b_{\text{fid}} = 2$ ) theory template, with a free amplitude fit to the data ( $A = b/b_{\text{fid}}$ ), is also indicated with a dashed line. The green dashed theory curve is fit only over a restricted analysis range (shaded region for scales  $100 < L < 1000$ ); the black solid curve is fit over the full  $L$  range shown in this plot. (The bandpowers are nearly independent, with the off-diagonal elements of the covariance matrix showing correlations of less than 13%). We find good consistency in both cases with the *Planck*-cosmology derived theory template.

and an excess on small scales. However, the difference was not found to deviate from zero by a statistically significant amount, with a  $\chi^2$  probability to exceed (PTE) of 0.29 (for the cosmology range). Nevertheless, we note that the difference is a good fit to a simplified foreground bias model (given by a 10% deficit in the cross-correlation at  $L < 800$ ); the  $\chi^2$  to this model is lower than for a fit to null by  $\Delta\chi^2 = 2.2$ .

Although the tSZ-free measurement contains no bias from tSZ, the measurement errors on large scales are similar, which highlights the power of this new technique in providing unbiased measurements that do not sacrifice significant signal-to-noise.<sup>16</sup>

<sup>16</sup>The fact that measurement uncertainties do not significantly increase in our method, although it removes foregrounds, is not just due to the inclusion of *Planck* data; indeed, a naive application of the standard quadratic estimator  $T_{\text{no-tSZ}} \vec{\nabla} T_{\text{no-tSZ}}$  to tSZ-deprojected ACT+*Planck* maps gives cross-correlation uncertainties that are  $\approx 50\%$  larger. *Planck* enables better multifrequency cleaning, rather than adding much raw statistical weight to the ACT maps.

We adopt the tSZ-cleaned cross-correlation as our standard analysis. We fit the cross-correlation with a fiducial theory model; this model uses both fiducial *Planck* parameters as well as a fiducial linear bias of  $b = 2$ , motivated by previous BOSS analyses [15]. The cross-correlation measurement as well as a fit of the amplitude of this fiducial model are shown in Figure (3.5). It can be seen that, for both the restricted analysis multipole range and the full range, the amplitudes obtained are consistent with the fiducial value ( $A = 1$ ). In particular, we obtain  $A = 0.92 \pm 0.12$  for a fit to the restricted analysis range and  $A = 1.02 \pm 0.10$  for the fit to the full range of scales. Both theory curves are a good fit to the measurements, with  $\chi^2$  PTEs of 0.25 and 0.28 respectively. Thus, we find good consistency in both cases with the *Planck*-cosmology derived theory template.

### 3.6 Systematics and Validation of the cross-correlation measurement

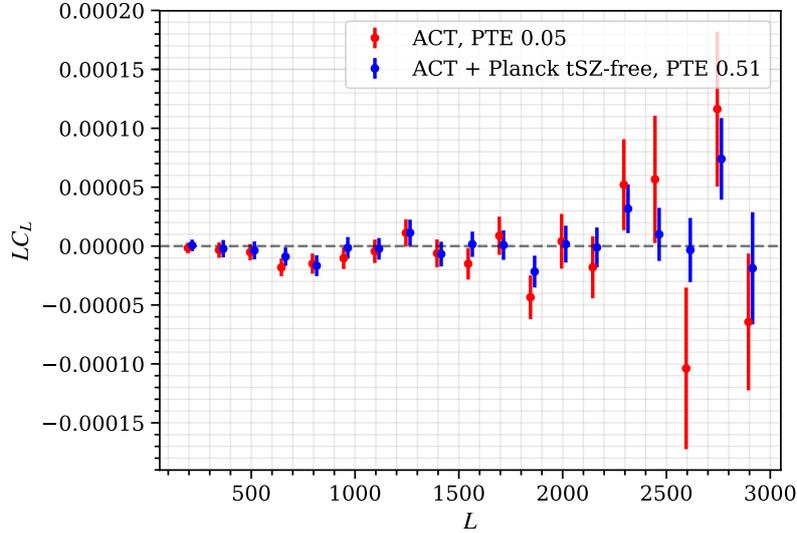
We perform several tests for systematic errors to validate both our lensing maps and our cross-correlation measurement. Note that the relevant covariance matrices are obtained from Monte Carlo simulations of each test. These covariances are used to derive a  $\chi^2$ -to-null probability to exceed (PTE) for every test.

Our first null test relies on the fact that we expect the cosmological lensing signal from gravitational scalar perturbations to give rise to gradient-like deflections. Hence, this deflection field should be irrotational, with zero curl.<sup>17</sup> In contrast, systematics that mimic lensing can have non-zero curl. Therefore, a detection of a curl signal can be a signature of unknown systematic errors present in our data. By using a quadratic estimator  $\hat{\Omega}^{XY}(\vec{L})$  similar to that for the lensing potential but with different filters [66] (essentially the dot product in the potential estimator is replaced by a cross product), it is possible to extract the curl signal and cross correlate it with the BOSS galaxy field. As shown in Figure (3.6), this cross-correlation signal is consistent with zero, with a PTE of 0.51 for the tSZ-cleaned lensing cross-correlation. We note that for the ACT-only cross-correlation, the PTE is only 0.05, although this may simply be due to a statistical fluctuation.

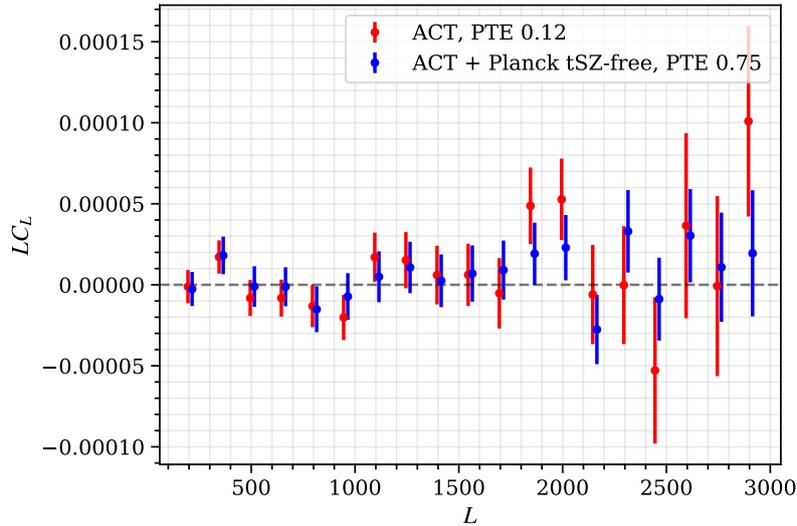
As a second test, we cross-correlate the galaxy map of one patch with the lensing convergence map of the other patch<sup>18</sup> and check for consistency with zero. It is very

<sup>17</sup>The potential cosmological curl signal coming from tensor perturbations at linear order or from scalar perturbations at second order is well below current sensitivity.

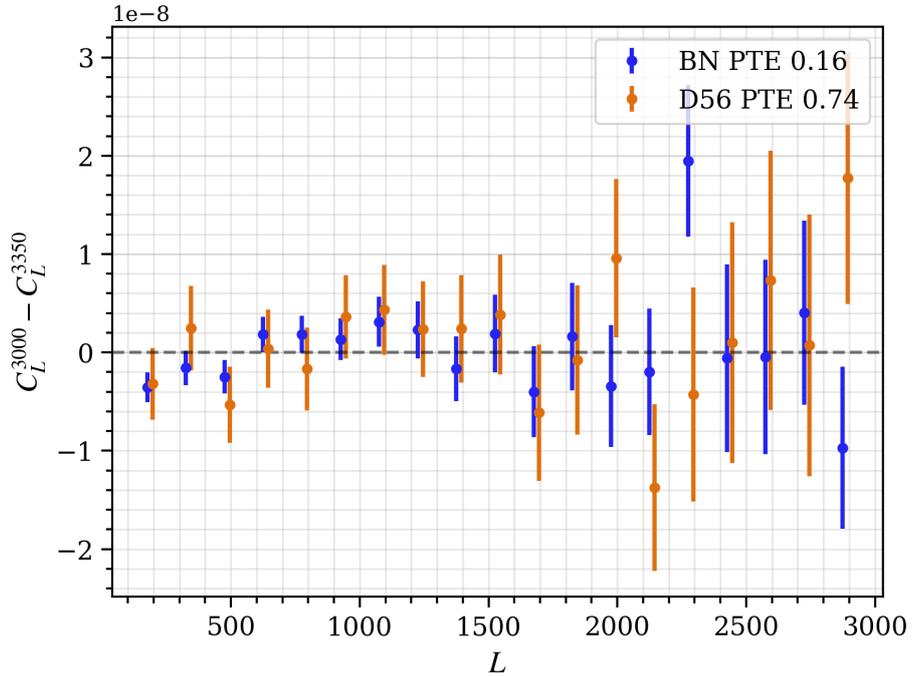
<sup>18</sup>To perform this correlation, we extend with zero values the maps of the smaller patch, in this case D56, so that the two fields have the same size.



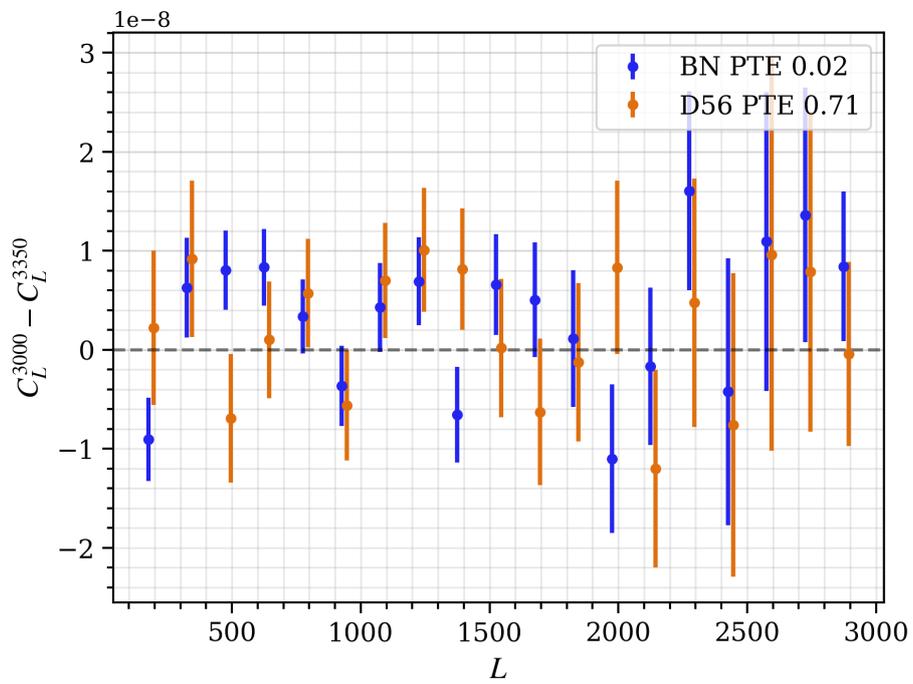
*Fig. 3.6* A curl null test: verification that the extracted curl-lensing – galaxy cross-correlation, which should be negligibly small in the absence of systematic errors, is consistent with the null hypothesis. The results shown are for a combination of both D56 and BN patches. The  $\chi^2$  probability-to-exceed (PTE) for this null test is also shown in the legend.



*Fig. 3.7* A null test verifying that cross-correlating the lensing map on one field with the galaxy map on the other field (and combining both spectra) gives a signal consistent with zero. Red points show ACT-only results, blue points show tSZ-deprojected lensing results. The fact that the PTEs in both cases are consistent with zero signal supports the conclusion that our uncertainty calculations are correct.



*Fig. 3.8* An extragalactic foreground null-test for the cleaned maps. We show the difference between the cross-correlations of CMASS with the tSZ-deprojected lensing maps for the cases of  $l_{\text{CMB,max}} = 3000$  and  $l_{\text{CMB,max}} = 3350$ , where  $l_{\text{CMB,max}}$  is the maximum CMB multipole used in the lensing reconstruction. Since extragalactic foregrounds rise rapidly towards high  $l$ , a substantial foreground residual in the cross-correlation would cause a null-test failure. However, our null-test results shown here are consistent with zero contamination for both fields (blue: points for BN, orange: points for D56).



*Fig. 3.9* The same test as shown in Figure (3.8), but applied to the ACT-only maps which have not been foreground cleaned (blue: points for BN, orange: points for D56). The PTE for BN shows a (mild) failure of the null test, as is expected if foreground residuals are important.

difficult to imagine systematics that would correlate fields that are so far apart, and so this test primarily serves as a validation of our covariance matrix and uncertainty calculation. In Figure (3.7) we see that the results of this null test are consistent with zero, with a PTE of 0.75 obtained for the tSZ-cleaned lensing map and 0.12 for the ACT-only map.

Thirdly, we wish to test for the presence of residual foreground-induced bias in the cross-correlation measurement, even though we expect to be insensitive to the dominant tSZ contamination when using our symmetric cleaned lensing estimator. To test for residual foreground biases from the CIB, kSZ [e.g. 89] or other sources (including those arising from incomplete tSZ cleaning), we make use of the fact that foreground contamination should become worse as the maximum CMB multipole  $\ell_{\text{CMB,max}}$  used in the lensing reconstruction increases. If our foreground cleaning is working as expected and residual foregrounds are negligible, results with a high  $\ell_{\text{CMB,max,high}}$  and a lower  $\ell_{\text{CMB,max,low}}$  used in the reconstruction should be consistent. In Figure (3.8) we show this foreground null test for the symmetric cleaned estimator; in particular, we plot the difference  $C_L^{\kappa_{\text{low}}^g} - C_L^{\kappa_{\text{high}}^g}$  of the cross-correlation  $C_L^{\kappa_{\text{high}}^g}$  with a higher lensing reconstruction  $\ell_{\text{CMB,max,high}} = 3350$  (the baseline used in this work) and the cross-correlation with a lower  $\ell_{\text{CMB,max,high}} = 3000$ ,  $C_L^{\kappa_{\text{low}}^g}$ . It can be seen that this difference is consistent with zero overall, with PTEs of 0.74 and 0.16 found for D56 and BN respectively. The error bars are obtained from simulations and hence take into account the covariance between the two spectra. For comparison, in Figure (3.9) we perform the same test for the ACT-only maps which are not free of tSZ; perhaps unsurprisingly, we find a (mild) null test failure (PTE of 0.02) for the BN patch, although the D56 PTE of 0.71 still appears acceptable.<sup>19</sup>

Finally, to check for sensitivity to large-scale systematics, we vary the lowest multipole  $L_{\text{min}}$  of the first bandpower of the cross-correlation measurement; we find that the value of the first bandpower is stable. This was the only null test done after we unblinded.

Our suite of null tests does not show evidence for foreground or systematic contamination to our measurement, as long as we use the symmetric cleaned lensing estimator. In particular, for the combined BN+D56 cleaned measurement we find a PTE of 0.28 for the foreground residual test, showing no evidence for foreground contamination in the cross-correlation.

<sup>19</sup>The fact that only one patch shows a null test failure does not have a clear explanation, although it may reflect the fact that our measurement errors are still fairly large compared to the foreground biases (and so fluctuations can be expected).

## 3.7 Discussion

In this chapter, we present maps of CMB lensing convergence derived from ACT observations made in 2014–15. The lensing maps are constructed in two different ways: first, by applying the standard quadratic lensing estimator to only ACTPol CMB data; second, by implementing a new “symmetric” foreground-cleaned lensing estimator, which makes use of component separated ACTPol+*Planck* CMB maps to return lensing maps that are free of tSZ-bias in cross-correlation.

We report combined cross-correlation measurements of our CMB lensing maps with BOSS CMASS galaxies at  $\approx 10\sigma$  significance. We find that the use of our new tSZ-free estimator does not significantly increase the size of measurement uncertainties.

We will release these lensing maps to enable other cross-correlation analyses with large-scale-structure. However, several caveats should be kept in mind when making use of these maps. Only the bispectrum  $\langle gT_{\text{tSZ}}T_{\text{tSZ}} \rangle$  tSZ contamination is nulled in our procedure, where  $T_{\text{tSZ}}$  is the tSZ signal and  $g$  is the large-scale structure field (e.g., galaxy overdensity or galaxy shear); this is the dominant source of contamination for near-term cross-correlations with  $z < 1$  structure. Users of these maps should be aware that high-redshift cross-correlations can be contaminated with the CIB field  $T_{\text{CIB}}$ , both through  $\langle gT_{\text{CIB}}T_{\text{CIB}} \rangle$  as well as through its correlation with the tSZ  $\langle gT_{\text{tSZ}}T_{\text{CIB}} \rangle$ . For cross-correlations where CIB contamination is more of a concern than tSZ contamination (e.g., for cross-correlations with the CIB itself), our pipeline allows the application of the analog of our symmetric cleaned estimator on CIB-deprojected maps from [170]. Such analyses should be validated on realistic simulations [e.g., 245, 267] to verify that the tSZ contamination is sub-dominant. Looking beyond the 2014 and 2015 data used in this work, high-resolution 230 GHz data collected with the Advanced ACTPol instrument from 2016 and onward should allow for simultaneous deprojection of both the tSZ and CIB contamination for use in symmetric cleaned estimators that are robust at all redshifts. The contamination from the kSZ will, however, remain, since the kSZ has the same blackbody frequency spectrum as the primary CMB, although the contamination is much lower in amplitude [75, 89]. Alternatives to our method include shear-only reconstruction [228] (which requires the inclusion of smaller scales in the CMB map to achieve similar signal-to-noise) and source hardening [202] (primarily targeted at reducing contamination from point sources and clusters). The optimal combination of all of these methods that minimizes bias (both from foregrounds and higher-order effects) and maximizes signal-to-noise will be discussed in Chapter 4.

We also caution users that the auto-spectrum of the lensing potential presents a much broader set of analysis challenges, both for mitigation of foregrounds (where the CIB contamination is expected to be larger [277]) and for characterization and subtraction of reconstruction noise bias. The latter requires an extensive set of simulations [e.g., 256, 270] and methods robust to mismatch of simulations and the observed sky [e.g., 195]. The CMB lensing auto-spectrum from ACT data from 2014 and 2015 will appear in a separate work. In addition, care should be taken when attempting to interpret the signal from stacking massive clusters on our released CMB lensing maps; first, because inpainting and masking steps can introduce complications, and second, because higher-order effects can bias the standard quadratic estimator near the most massive clusters [125].

This work lays the foundation for upcoming, higher precision ACTPol and Advanced ACT cross-correlations with galaxy and lensing surveys. For upcoming cross-correlation analyses with ACT and other experiments, powerful methods to obtain foreground free measurements are necessary; our work represents one promising solution to this problem.

## Optimizing foreground mitigation for CMB lensing with combined multifrequency and geometric methods

### Summary

*A key challenge for current and upcoming CMB lensing measurements is their sensitivity to biases from extragalactic foregrounds, such as Sunyaev-Zel'dovich (SZ) signals or cosmic infrared background emission. Several methods have been developed to mitigate these lensing foreground biases, dividing broadly into multi-frequency cleaning approaches and modifications to the estimator geometry, but how to optimally combine these methods has not yet been explored in detail. In this chapter, we examine which combination of lensing foreground mitigation strategies is best able to reduce the impact of foreground contamination for a Simons-Observatory-like experiment while preserving maximal signal-to-noise. Although the optimal combination obtained depends on whether bias- or variance-reduction are prioritized and on whether polarization data is used, generally, we find that combinations involving both geometric (profile hardening, source hardening or shear) and multifrequency (symmetric cleaning) methods perform best. For lensing power spectrum measurements from temperature (polarization and temperature), our combined estimator methods are able to reduce the bias below  $\sigma/4$  or 0.3% (0.1%), a factor of 16 (30) lower than the standard QE bias, at a modest signal-to-noise cost of only 18% (12%). In contrast, single-method foreground-mitigation approaches struggle to reduce the bias to a negligible level below  $\sigma/2$  without incurring a large noise penalty. For upcoming and current experiments, our combined methods therefore represent a promising approach for making lensing measurements with negligible foreground bias.*

## 4.1 Introduction

Along their paths to our telescopes, photons of the cosmic microwave background radiation (CMB) are deflected by the gravitational influence of matter in our Universe. This leads to a remapping of CMB photons that depends on a weighted integral of the matter perturbations along the line of sight. The ability to map directly the projected mass distribution out to high redshifts makes CMB lensing a powerful source of cosmological information (such as constraints on neutrino masses or dark energy properties) (e.g. [219, 5]).

When reconstructing the projected matter fluctuations from observations, CMB foreground contamination in temperature is expected to induce  $\sim 5\%$ -level biases in the CMB lensing power spectrum and lensing cross-correlations with tracers of the matter field [202, 278, 89, 36, 263, 75, 169, 229, 226]. These biases are especially concerning for high-resolution ground-based CMB experiments, such as AdvACT, SPT-3G and Simons Observatory, since these experiments still rely heavily on lensing reconstruction from temperature and derive more information from small angular scales which have higher levels of foreground contamination.

Several mitigation strategies have been proposed to address this key foreground challenge for CMB lensing. Broadly, these strategies divide into geometric methods and frequency-based methods. Geometric methods [196, 229, 226], which include bias hardening, profile hardening, and shear reconstruction, aim to modify the lensing estimators' weight functions in order to null or reduce the biases induced by foregrounds. Frequency-based methods [77, 224, 169, 2, 171, 14, 43], which include foreground deprojection in a constrained or partially constrained ILC, as well as gradient cleaning and symmetric cleaning, use the departure of the foregrounds' spectral properties from a blackbody to null or reduce the foreground bias levels.

While many of these methods can reduce foreground biases quite effectively, with increasing measurement precision the requirements on bias mitigation are becoming increasingly stringent, and further improvements are becoming well motivated. A key question for upcoming experiments therefore is: which combination of bias mitigation strategies minimizes the biases most effectively while preserving as much signal-to-noise as possible?

To address this question, we consider estimators composing – i.e., simultaneously applying – geometric and frequency-based mitigation methods; we also consider linear combinations of different estimators at the lensing reconstruction level and optimise this linear combination to best mitigate foreground-induced lensing biases while maximising the signal-to-noise.

The remainder of this chapter is structured as follows. In Section 4.2, we introduce our method for combining different estimators and mitigation approaches. In 4.3, we define bias and noise measures for the combined estimator. In 4.4 we present the optimization formalism, followed by our results and their discussion in 4.5. We conclude in Section 4.6.

## 4.2 CMB Lensing foreground biases for estimators

### 4.2.1 Quadratic CMB Lensing Estimators

Weak gravitational lensing of the CMB induces correlations between different modes of the CMB, because a fixed lensing field breaks the statistical isotropy of the primordial CMB. One can use these off-diagonal correlations to write a quadratic estimator for the CMB lensing convergence field from the observed beam-deconvolved lensed CMB temperature field  $T$ . This estimator has the form [130]<sup>1</sup>

$$\hat{\kappa}(\vec{L}) = \int_{\vec{\ell}} T(\vec{\ell}) T(\vec{L} - \vec{\ell}) g(\vec{\ell}, \vec{L} - \vec{\ell}), \quad (4.1)$$

where

$$\int_{\vec{\ell}} \equiv \int \frac{d^2 \vec{\ell}}{(2\pi)^2}, \quad (4.2)$$

and  $g$  is a weight-function that satisfies the unit response condition to the true CMB lensing convergence field,  $\langle \hat{\kappa} \rangle_{\text{CMB}} = \kappa$ , where  $\langle \dots \rangle_{\text{CMB}}$  denotes averaging over the primordial CMB while fixing the lensing mode to be reconstructed. This, in turn implies that

$$\int_{\vec{\ell}} f(\vec{\ell}, \vec{L} - \vec{\ell}) g(\vec{\ell}, \vec{L} - \vec{\ell}) = 1, \quad (4.3)$$

where  $f$  is a response function that encodes the response of the off-diagonal CMB temperature two-point correlation function to lensing,  $\langle T(\vec{\ell}) T(\vec{L} - \vec{\ell}) \rangle_{\text{CMB}} = f(\vec{\ell}, \vec{L} - \vec{\ell}) \kappa(\vec{L})$ .

### 4.2.2 Origin of foreground biases

The observed CMB temperature field  $T$  contains contributions from both the lensed CMB  $T_{\text{CMB}}$  and foregrounds  $T_f$ , so that  $T = T_{\text{CMB}} + T_f$ , where here we ignore any noise contribution. The foregrounds, which are non-Gaussian and also correlated with

<sup>1</sup>We focus on temperature-only reconstruction for now, and later we will discuss including polarization.

the lensing convergence, give rise to biases in both the auto-correlation of a CMB lensing quadratic estimator  $\hat{Q}[T, T]$ , and to the cross-correlation of  $\hat{Q}$  with an external matter tracer  $g_m$ . As discussed in [277, 202, 229], and explained in detail in Appendix (B.1), foregrounds induce three bias terms to the reconstructed auto-spectrum: a trispectrum term of the form  $T_f^4$ , and two bispectrum terms that involve  $T_f^2 \times \kappa$ . For a CMB lensing cross-spectrum with an LSS tracer, the bias appears as a bispectrum of the form  $\sim T_f^2 \times g_m$ . These biases can easily lead to an incorrect inference in parameters, e.g., the amplitude of matter fluctuations, if not treated properly.

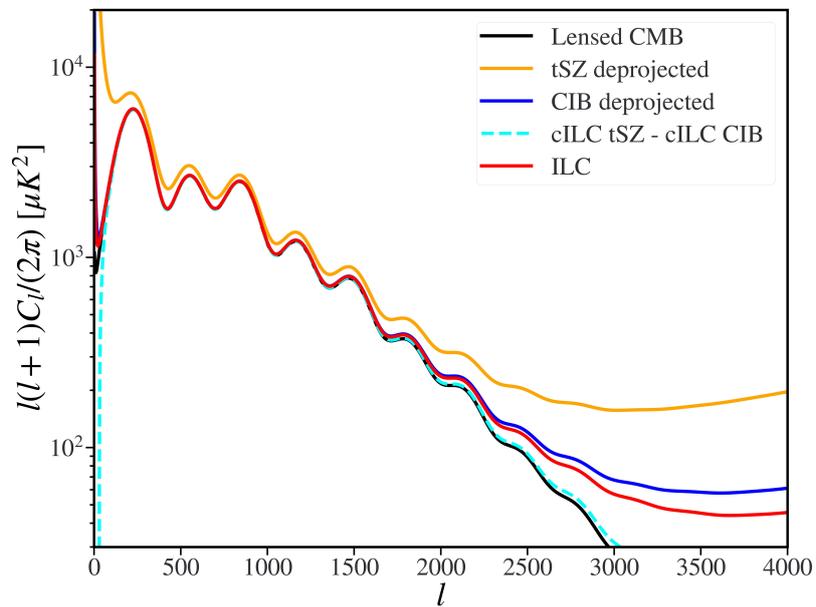
### 4.2.3 Current foreground mitigation methods and estimators

Several modifications to CMB lensing estimators have been proposed to extract an unbiased CMB lensing signal; they differ both in the degree to which they mitigate foreground biases and to which they lose signal-to-noise. In this chapter we will consider the following foreground-mitigating CMB lensing estimators, some of which rely on geometric mitigation of the estimator kernels, and others on multi-frequency methods:

- The standard Quadratic Estimator [130] (QE): this estimator has the smallest possible variance. On the other hand, it is not immune to the biases induced by foregrounds due to the contamination of the temperature maps.
- Bias hardened estimators [196, 226], such as point source-hardening (PSH) and profile hardening (PH), which modify the estimator weights to null a component of known response at the cost of increasing the variance.
- The shear estimator [229] (SH), which isolates the quadrupole (shear) component of the lensing response  $f$ , and is generally insensitive to foregrounds with symmetric profiles about the line of sight.
- The symmetric multi-frequency cleaned estimator [169] and chapter 3 (Symm), which allows for the nulling of some foreground effects using multi-frequency cleaned data, but has been constructed to reduce the noise cost incurred.

In this chapter we use a tSZ-like profile for the PH estimator, identical to the one considered in [226], i.e. the square root of a fiducial tSZ power spectrum on small scales.

In Figure 4.1 we plot the temperature power spectra of several multi-frequency combinations for an SO-like experiment, while in Figure 4.2 we show the lensing reconstruction noise for different estimators, assuming  $\ell_{\max} = 3500$  is used in the



*Fig. 4.1* The temperature power spectra of different multi-frequency combinations of frequency maps for a Simons Observatory-like experiment: in red is the standard internal linear combination (ILC); in blue is the ILC with CIB deprojection; in orange is the ILC with tSZ deprojection. In cyan we show the cross correlation between the tSZ-deprojected and CIB-deprojected combinations. In black is the lensed CMB theory curve. For the tSZ-deprojected curve in orange, the large power on large scales is due to atmospheric noise.

reconstruction, as well as the total foreground biases for a few estimators (see Appendix (B.7), Figure B.13 for the biases of the remaining estimators).

## 4.3 Optimal combination of CMB lensing estimators: formalism

### 4.3.1 Amplitude shift on the CMB lensing power

For our investigations we must be able to describe appropriately the impact of foregrounds on the measured CMB lensing power spectrum. For simplicity we will focus on constraints on the amplitude  $A$  of the lensing power spectrum or the cross-correlation of the lensing convergence with a matter tracer – we expect this parameter to be a good proxy for the most relevant applications of the lensing power spectrum, such as measuring  $\sigma_8$  or the neutrino mass. Effectively we have  $\hat{A}_L^\alpha \propto \hat{C}_L^{\kappa\alpha}$ , with  $\alpha = \kappa$  or  $g$ , where  $g$  is a tracer of the underlying matter distribution (e.g. galaxies). Therefore, to lowest order, a bias in the measured power spectrum leads to a bias to the inferred amplitude, i.e.  $\delta A_L^\alpha \propto \delta C_L^{\kappa\alpha}$ . Let us define the measured lensing amplitude per mode  $\vec{L}$  with respect to a fiducial cosmology:

$$\hat{A}(\vec{L}) = \hat{C}_L^{\kappa\alpha} / C_L^{\kappa\alpha, \text{fid}}, \quad (4.4)$$

where we have suppressed the explicit dependence of  $\hat{A}$  on  $\alpha$ . Assuming that  $\kappa$  and  $g_m$  are Gaussian fields, for one single mode the variance of this estimator is related to the variance of the power spectrum estimator  $\hat{C}_L^{\kappa\alpha}$

$$\sigma_L^2 = (C_L^{\kappa\alpha} + N_L^{\kappa\alpha})^2 + (C_L^{\kappa\kappa} + N_L^{\kappa\kappa})(C_L^{\alpha\alpha} + N_L^{\alpha\alpha}), \quad (4.5)$$

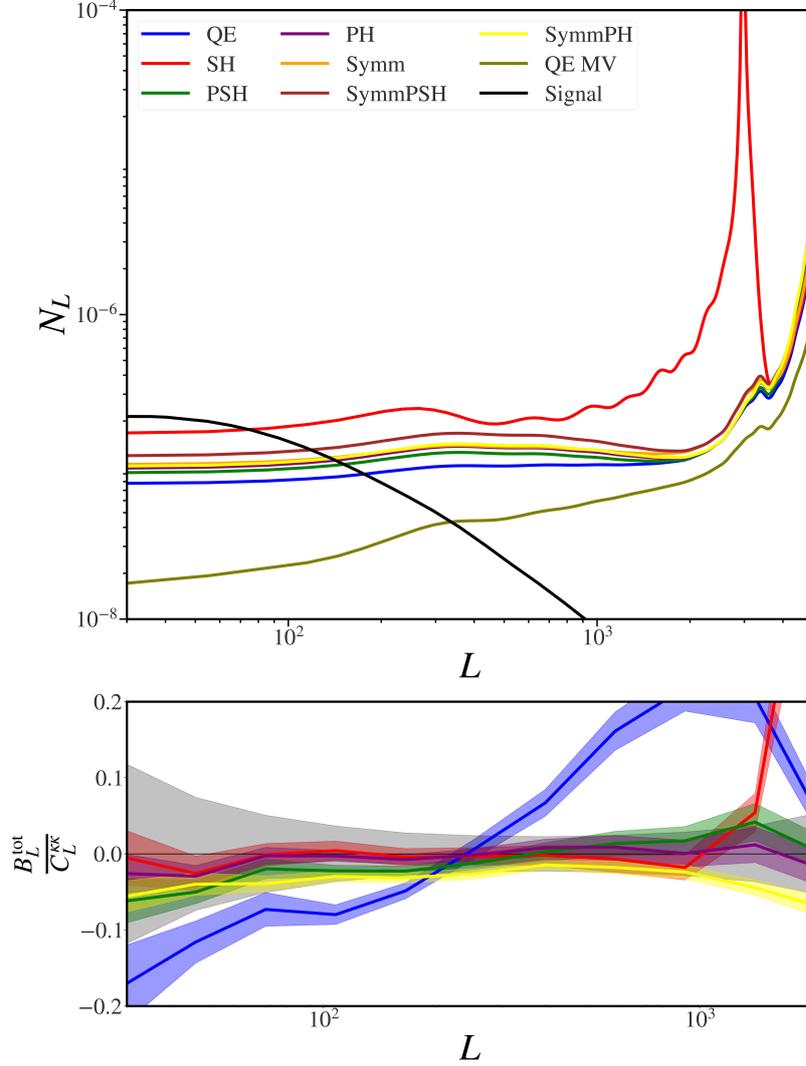
where  $N_L^{\alpha\beta}$  a noise component, equal to the lensing reconstruction noise if  $\alpha = \beta = \kappa$ , the Poisson shot noise if  $\alpha = \beta = g_m$ , and zero if  $\alpha \neq \beta$ . The bias to the estimator  $\hat{A}(\vec{L})$  is simply

$$\mathcal{B}(\hat{A}(\vec{L})) = \delta C_L^{\kappa\alpha} / C_L^{\kappa\alpha, \text{fid}}. \quad (4.6)$$

Given a range of modes, we can combine the measurements of  $A$  at each  $\vec{L}$  to obtain a global estimator:

$$\hat{A} = \int_{\vec{L}} w(\vec{L}) \hat{A}(\vec{L}), \quad (4.7)$$

where the weights  $w$  satisfy  $\int_{\vec{L}} w(\vec{L}) = 1$  to ensure that the estimator  $\hat{A}$  is unbiased to the fiducial cosmology ( $\langle \hat{A} \rangle = A_{\text{fid}} = 1$ ). When foregrounds are present, the estimator



*Fig. 4.2* CMB lensing reconstruction noise curves for the standard quadratic estimator (QE), shear (SH), point source hardened estimator (PSH), and the profile hardened estimator with a tSZ profile (PH), all applied to a minimum variance temperature-only ILC map, as well as noise curves for the standard symmetric estimator with tSZ-deprojection (Symm) and two new estimators that we introduce later in the text: point source hardening on the symmetric estimator with tSZ-deprojection (SymmPSH), profile hardening on the symmetric estimator with CIB-deprojection (SymmPH), and finally the minimum variance estimator with both temperature and polarization (QE MV). For the temperature maps we take  $\ell_{\min}, \ell_{\max} = 30, 3500$  in this figure, while for polarization we assume  $\ell_{\min} = 30, \ell_{\max, \text{pol}} = 5000$ . In black we show the CMB lensing signal  $C_L^{\kappa\kappa}$ . In the bottom panel we show the total bias to the CMB lensing autospectrum (as a fraction of the signal power) for some of the estimators. The relative statistical error for the standard QE is shown in grey. Note that the standard QE is biased at most scales.

$\hat{A}$  acquires a bias:

$$b(\hat{A}) = \int_{\vec{L}} w(\vec{L}) \mathcal{B}(\hat{A}(\vec{L})). \quad (4.8)$$

The variance of the estimator  $\hat{A}$  is given by

$$\sigma^2(\hat{A}) = \frac{1}{4\pi f_{\text{sky}}} \int_{\vec{L}} w_L^2 \frac{\sigma_L^2}{(C_L^{\kappa\alpha, \text{fid}})^2}, \quad (4.9)$$

where  $f_{\text{sky}}$  is the covered sky fraction, and  $w_L \equiv w(\vec{L})$ . Of particular relevance is the minimum-variance estimator, which has weights

$$w_L^{\text{MV}} = \frac{(C_L^{\kappa\alpha, \text{fid}})^2}{\sigma_L^2} \left[ \int_{\vec{L}} \frac{(C_L^{\kappa\alpha, \text{fid}})^2}{\sigma_L^2} \right]^{-1}. \quad (4.10)$$

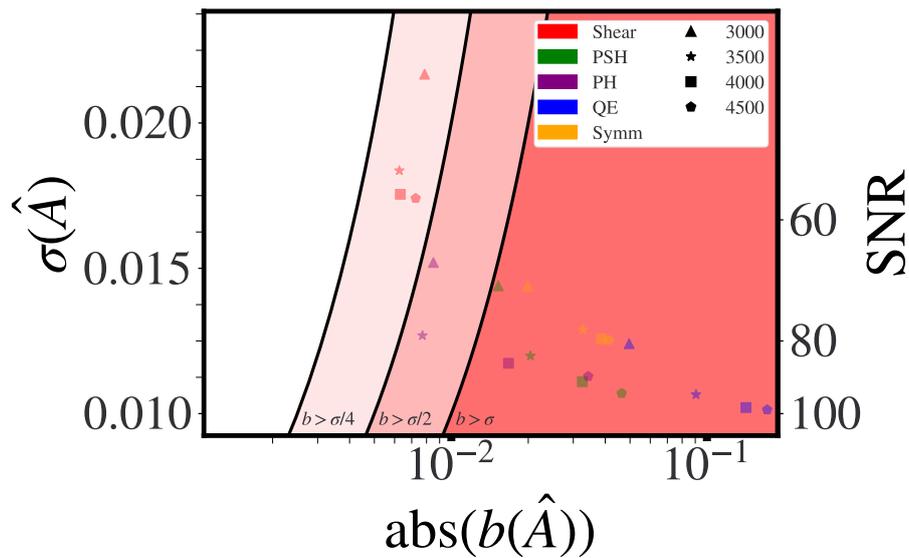
In Figure 4.3 we plot the CMB lensing error and bias for the foreground-mitigating estimators proposed thus far, assuming several different values for the reconstruction  $l_{\text{max,TT}}$  and summing over CMB lensing modes  $30 \leq |\vec{L}| \leq 1200$ .

In calculating the bias, we choose to replace the “true” biases  $B_L^i$  for the estimator  $i$ , as measured from a set of simulations described in section 4.5, with a smoothed version of the absolute value of the biases. This choice will be justified later in section 4.4.

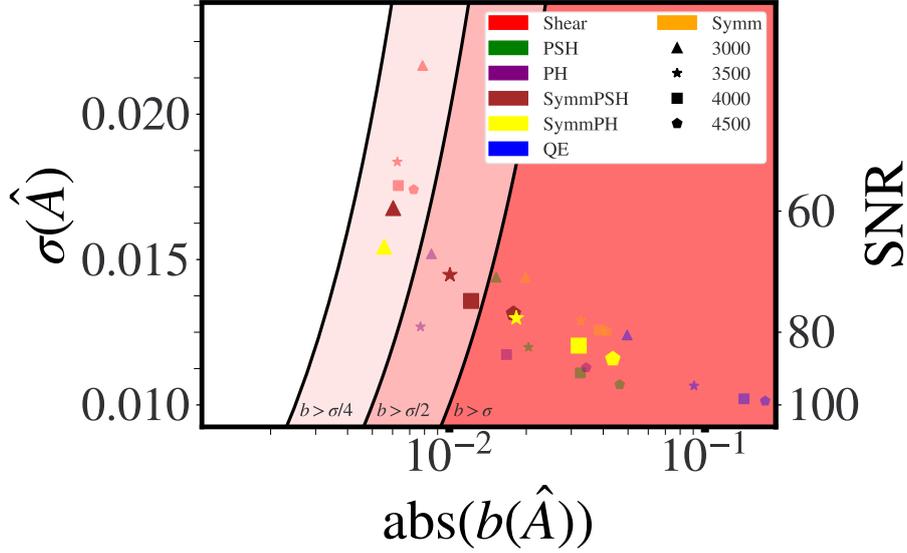
It is clear that in most cases we are in a regime where bias is not negligible. For the estimators which achieve  $b < \sigma$  we are in general paying a substantial noise penalty, even for profile hardening (PH) at  $l_{\text{max,TT}} = 3500$ . Furthermore, to ensure that bias is truly subdominant and negligible (despite uncertainties in simulations and modeling), we would like to be in a regime where bias is well below the statistical uncertainty, such as  $b \leq \sigma/4$  – and this cannot be realistically achieved with the currently proposed estimators.

### 4.3.2 Amplitude shift on CMB lensing power from composing geometrical and multi-frequency methods

To improve upon the current situation, we first propose two new CMB lensing estimators; these are constructed by simultaneously applying (or “composing”) both multi-frequency cleaning and the bias hardening operation in an estimator. In particular, first we apply the symmetric estimator (which involves multifrequency cleaning), then we apply bias hardening on the symmetric estimator as explained in Appendix (B.2). For the first new estimator, we consider a symmetric estimator with a tSZ-deprojected map, and then apply point source hardening. For the second new estimator,



*Fig. 4.3* Values of measurement standard deviation  $\sigma$  and bias  $b$  for the CMB lensing power spectrum amplitude  $A$ , assuming only temperature ( $TT$ ) reconstruction is used in the lensing power spectrum measurement. Results are shown for the different estimators with different types of foreground mitigation that have been proposed so far; several values are assumed and plotted for the maximum CMB multipole used in the lensing reconstruction. In this and the following plots, [QE, SH, PSH, PH, Symm] indicate the standard quadratic estimator, shear estimator, point source hardened estimator, profile hardened estimator, and symmetric tSZ-deprojected estimator respectively. The colored bands represent regimes where the bias is greater than a certain fraction of the statistical noise. We can see clearly that a trade-off between bias and noise exists for the different estimators proposed thus far: estimators that aim to reduce biases unfortunately incur a significant noise penalty. We would like to be able to reduce both foreground bias and variance of lensing estimators simultaneously (obtaining new estimators that are somewhat closer to the lower left corner of this plot).



*Fig. 4.4* As for Figure 4.3, but with the addition of two new combined estimators: i) SymmPSH, which stands for point source hardening applied to a symmetric estimator with tSZ deprojected; ii) SymmPH, given by profile hardening applied to a symmetric estimator with CIB deprojected. We can see that, for some configurations, these new estimators allow us to obtain somewhat lower bias at the same noise in comparison with the existing estimators shown in Figure 4.3, although potential remains for further improvements.

we consider using a symmetric estimator with a CIB-deprojected map (see [171] for the first implementation of CIB deprojection on real CMB data), and then apply profile hardening. The logic behind these choices is that we employ one mitigation method that targets the tSZ (tSZ deprojection or profile hardening) alongside another that focuses on the CIB (CIB deprojection or point source hardening). We refer to these two new estimators as SymmPSH (using tSZ deprojection with point source hardening), and SymmPH (using CIB deprojection with profile hardening), respectively.

We can see from Figure 4.4 that these estimators perform well: they generally have lower biases given the same variance as the previously proposed estimators and have  $b < \sigma$ . However, further improvements are still motivated, since we have not yet found that biases are completely negligible such that  $b \ll \sigma$ . For this purpose, we will now investigate linear combinations of different estimators  $\hat{Q}_i$ .

### 4.3.3 Amplitude shift on CMB lensing power for combined estimators

Using the same definitions as above, we can straightforwardly derive the bias and variance of the lensing amplitude estimator  $\hat{A}$  when computed from a linear combination

of different individual quadratic estimators (with different foreground mitigation methods)  $\hat{Q}_i$ . We can define this linear combination as follows:

$$\hat{Q}(\vec{L}) \equiv \sum_i a_i(L) \hat{Q}_i(\vec{L}), \quad (4.11)$$

where  $\sum_i a_i(L) = 1$  to ensure unit response to the CMB:  $\langle \hat{Q} \rangle_{\text{CMB}} = \kappa_{\text{CMB}}$ . For the estimator  $\hat{Q}$ , the measured power spectrum  $\hat{C}_L^{\kappa\alpha}$  takes the form:

$$\hat{C}_L^{\kappa\alpha} = \sum_{i,j} a_i(L) b_j(L) \hat{C}_L^{\kappa\alpha,ij}, \quad (4.12)$$

where  $\hat{C}_L^{\kappa\alpha,ij}$  is the reconstructed power spectrum from the correlation of the individual  $i, j$  estimators, and  $b_j(L) = a_j(L)$  when  $\alpha = \kappa$ , or  $b_j(L) = \delta_{j1}^K$  when  $\alpha = g$ .

The estimator for the amplitude now takes the form

$$\hat{A} = \int_{\vec{L}} w(\vec{L}) \sum_{i,j} a_i b_j \hat{C}_L^{\kappa\alpha,ij} / C_L^{\kappa\alpha,\text{fid}}. \quad (4.13)$$

In the presence of foregrounds, this estimator obtains a bias, which can be expressed as

$$b(\hat{A}) = \int_{\vec{L}} w(\vec{L}) \sum_{i,j} a_i b_j \frac{B_L^{ij}}{C_L^{\kappa\alpha,\text{fid}}}, \quad (4.14)$$

where  $B_L^{ij} = \delta C_L^{\kappa\alpha,ij}$  is the foreground-induced lensing bias from the cross-correlation of the individual  $i, j$  estimators. The variance (assuming the lensing reconstruction noise can be approximated as Gaussian) is given by

$$\sigma^2(\hat{A}) = \frac{1}{4\pi f_{\text{sky}}} \int_{\vec{L}} w_L^2 \sum_{ijmn} \frac{\Theta_L^{ijmn} a_i a_j b_m b_n}{(C_L^{\kappa\alpha,\text{fid}})^2}, \quad (4.15)$$

where we have suppressed the  $L$  dependence of  $a_i$  and  $b_i$ , and we have defined

$$\Theta_L^{ijmn} \equiv (C_L^{\kappa\alpha} + N_L^{\kappa\alpha,in})(C_L^{\kappa\alpha} + N_L^{\kappa\alpha,jm}) + (C_L^{\kappa\kappa} + N_L^{\kappa\kappa,im})(C_L^{\alpha\alpha} + N_L^{\alpha\alpha,jn}). \quad (4.16)$$

Again, when  $\alpha \neq \beta$ , the noise term  $N_L^{\alpha\beta,ij} = 0$ . When  $\alpha = \beta = g$ ,  $N_L^{gg,ij} \equiv N^{gg}$  is just the Poisson shot noise. When  $\alpha = \beta = \kappa$ ,  $N_L^{\alpha\beta,ij}$  is the lensing noise associated with the cross-correlation of the  $i, j$  estimators.

We note that while the variance  $\sigma^2(\hat{A})$  is guaranteed to decrease as more  $L$ s are summed over, the same is not true for the bias. Thus, even in the case where the bias  $B_L$  is subdominant to the noise  $\sqrt{\Theta_L}$  for each  $L$ , one can still have  $b^2(\hat{A}) > \sigma^2(\hat{A})$ .

## 4.4 Finding the optimal combination

Now that we have introduced all the relevant definitions, we can begin to address the key question of this chapter: what is the optimal linear combination for minimizing both foreground bias and measurement noise? To answer this question, we define a loss function, inspired by the mean squared error  $\langle(\hat{A} - A)^2\rangle$ , which allows us to parameterize our goals in this optimization: i.e. to what extent we prioritize bias or variance. To this end we consider the following loss function

$$\mathcal{L}[a_i, w; f_b] = \sigma^2(\hat{A}) + f_b^2 b^2(\hat{A}), \quad (4.17)$$

where  $\sigma^2(\hat{A})$  and  $b(\hat{A})$  are defined in Eqs. (4.14) and (4.15) respectively, and the  $f_b$  parameter regulates the importance of the bias in the total functional, with a higher  $f_b$  assigning a higher importance to bias reduction.<sup>2</sup> Equation (4.17) represents a general quartic optimization problem, which we numerically minimize by varying  $a_i(L)$  and  $w(\vec{L})$  subject to the constraints

$$\sum_i a_i(L) = \int_{\vec{L}} w(\vec{L}) = 1. \quad (4.18)$$

We additionally impose the constraint  $a_i(L) \geq 0$  on the coefficients, which make our optimized solutions less finely-tuned to the simulated biases.<sup>3</sup>

In our optimization we choose to replace the “true” biases  $B_L^{ij}$ , as measured from simulations, with a smoothed version of the absolute value of the biases. That is, when minimizing Eq. (4.17), we replace  $B_L^{ij}$  in Eq. (4.14) with  $|B_L^{ij}|_s > 0$ , where the subscript  $s$  denotes a smoothing operation, which is described in Appendix (B.5). Doing so avoids potential exact cancellations in bias among different  $L$ s in order to be conservative (and less simulation-dependent), as seen for example in Fig. 4.2. We discuss how this choice impacts our results in Appendix (B.5).

We note that the case  $f_b = 0$  corresponds to minimizing the variance  $\sigma^2(\hat{A})$ . Since the integrand in Eq. (4.15) is positive, minimizing the full variance corresponds to minimizing the integrand for each  $\vec{L}$ . In other words, the ideal solution corresponds

<sup>2</sup>One can introduce an extra function  $r$  too, monotonic in the weights  $a$  and biases matrix  $\mathbf{B}$ : this should be chosen to act as a prior on the weights, or as a regularizer. The main reason for such a term is to not overfit the specific set of simulations or theoretical models used for calculating the biases, as there is an inherent modeling uncertainty in these. A regularizer will be also useful as a smoother of the optimal solution, to have more stable, non-oscillating solutions.

<sup>3</sup>For example, if the biases of two estimators were known perfectly well, and if the coefficients weren’t restricted to be non-negative, one could easily solve for coefficients which cancelled these biases. This cancellation is simulation dependent, and could result in significant biases on real data if the biases to the individual estimators are non-negligible.

to the minimum variance combination of the individual estimators, and setting  $w(\vec{L}) = w_L^{\text{MV}}$  (i.e. Eq. (4.10) with  $\sum_{ijmn} \Theta_L^{ijmn} a_i a_j b_m b_n$  replacing  $\sigma_L^2$ ).

## 4.5 Optimisation in practice

### 4.5.1 Calculating foreground CMB lensing biases

In this chapter we consider a Simons Observatory-like experiment, with six observational frequency channels ([27, 39, 93, 145, 255, 280] GHz). We assume the goal noise levels [5] for the detector and atmospheric noise contributions.<sup>4</sup> We follow [84] when modeling the foreground power spectra and their SEDs, as implemented in the code LensQuEst.<sup>5</sup>

The foreground-induced CMB lensing biases that we consider here come from the total sum of extragalactic foregrounds, namely tSZ + CIB + kSZ + radio PS, as given by the non-Gaussian foreground simulations of [244] at 150 GHz. To create ILC combinations from these simulations we do the following: first, we mask the sum of extragalactic foregrounds with a mask obtained by inpainting [45] disks with a 3 arcmin radius around individual detected point sources with flux density higher than 5 mJy at 150 GHz, picking up also clusters. After masking, we Fourier transform the simulations, rescaling them from 150 GHz to the corresponding SO frequencies using the SEDs of [84],<sup>6</sup> and create the ILC/cILC combinations, with a total theory power spectrum given by lensed CMB added to experimental noise, galactic dust, and extragalactic foreground contributions from theory spectra, as given by [84]. Then we take the inverse Fourier transform and mask again with the same mask (equal to one outside the inpainted disks).

We then follow the method of [229, 226] to estimate the foreground-induced non-Gaussian CMB lensing biases and hence obtain the matrix  $B^{ij}$  for the bias arising from the cross of estimator  $i$  with estimator  $j$ . We describe the specific calculations in Appendix (B.1).

---

<sup>4</sup>The noise curves are calculated using the noise calculator V3\_calc available at a private repository on the SO Github.

<sup>5</sup><https://github.com/EmmanuelSchaan/LensQuEst>

<sup>6</sup>Note that we are implicitly assuming that CIB can be simply rescaled, but in reality there is a decoherence among CIB observed at different frequencies, i.e. the correlation among CIB maps at different frequencies is not one, although quite high in the simulations.

## 4.5.2 Results

In the following subsections we present our optimization results. Aside from  $f_b = 0$ , for all cases we optimize over the [SH, PSH, PH, SymmPSH, SymmPH] estimators as explained in Appendix (B.5) (recall that these abbreviations indicate shear, point source hardening, profile hardening, symmetric tSZ deprojection with point source hardening, symmetric CIB deprojection with profile hardening); we also allow the maximum multipole to vary separately for each estimator in our optimization, with the options  $l_{\max,TT} = [3000, 3500, 4000, 4500]$ .

### Lensing auto-spectrum

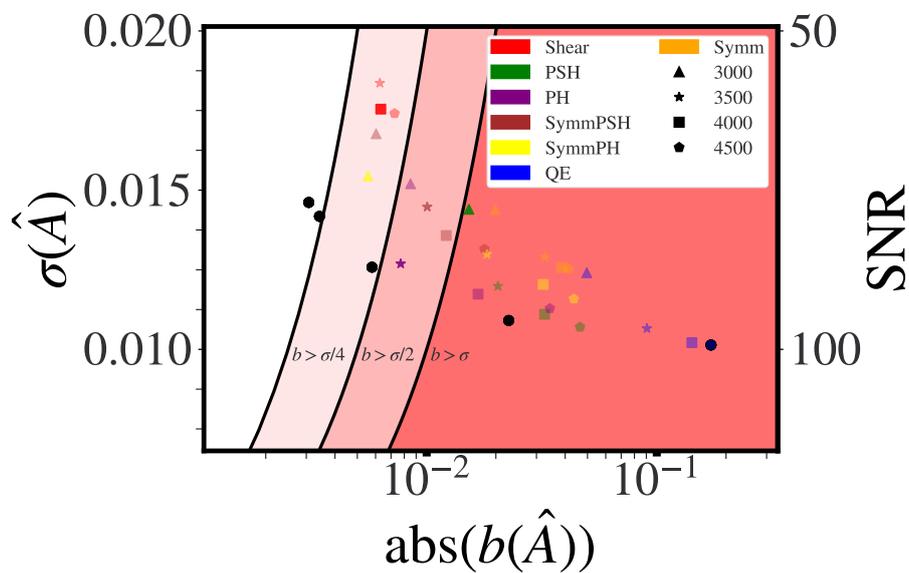
**Temperature-only lensing reconstruction** In Figure 4.5 we show our results for combining estimators using only temperature data; here solid black dots show the performance of our new optimal combinations (on the  $x$ -axis we again plot the absolute value of the induced foreground bias on the amplitude of the CMB lensing power spectrum, as described in Appendix (B.5). and on the  $y$ -axis the noise on the lensing amplitude). For black dots from right to left, we vary  $f_b$  from  $f_b = 0, 0.1, 1, 2, 4$ , progressively increasing the importance of the bias in the noise-bias trade-off. We can see that all the combined estimators perform significantly better than the single estimators, giving a lower bias for the same noise. Figure 4.6 shows in more detail the composition of these optimal combinations (each piechart in this figure describes the corresponding solid black dot in the previous Figure 4.5).

Our reference case for  $TT$  will be QE at  $l_{\max} = 3000$ , for which we have a bias of 5%; this estimator gives a total SNR of 80.

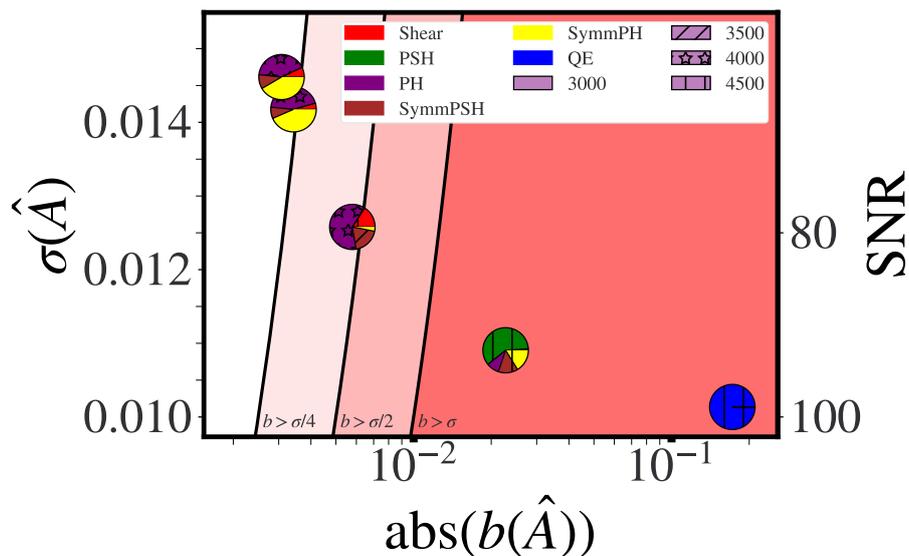
When  $f_b = 0$ , we have the minimum-variance solution, which is just the standard QE. The signal-to-noise of the temperature-only reconstruction in this case is close to 100 since  $l_{\max} = 4500$  now.

For  $f_b = 0.1$ , PSH with a high  $l_{\max} = 4500$  gains importance, although there is also some contribution from other estimators. In this case bias is nearly irrelevant in our optimization, so that the non-QE estimator with the highest signal-to-noise dominates. This combination decreases the bias by a factor of 2.2 with respect to QE at  $l_{\max} = 3000$  (although the bias is still quite large, at the level of a few percent); this reduction in bias still has a lower noise compared to the reference case of QE, with a SNR above 90.

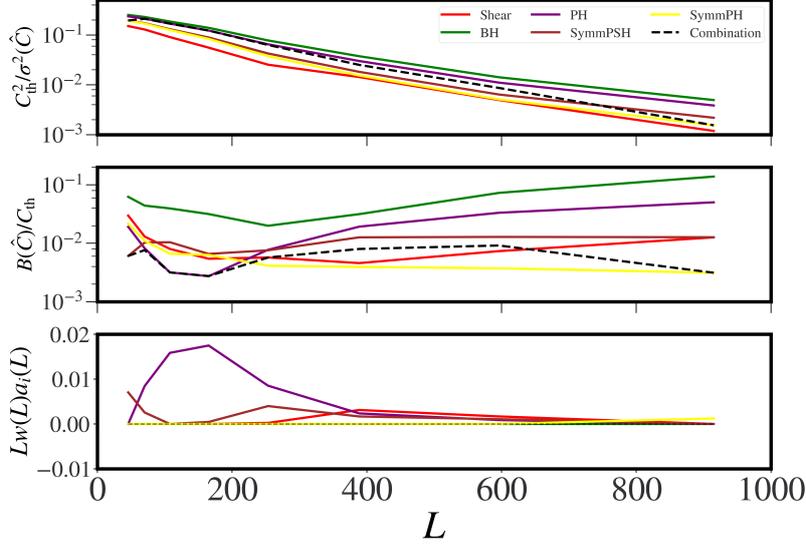
For  $f_b = 1$ , a case where we assign minimizing bias and variance similar priority, our solution decreases the bias by 8.6 times compared to QE at  $l_{\max} = 3000$ . The solution gives a bias of  $\approx 0.6\%$ , corresponding to  $\sigma/2$ , at a moderate noise cost similar to the simple QE, corresponding to an SNR of 80. In Figure 4.7 we illustrate more clearly



*Fig. 4.5* As for Figures 4.3 and 4.4, but now including results from performing an optimized linear combination of individual estimators (in this plot we only consider lensing measurement from temperature). The solid black dots show the optimized points, with results shown for optimization with  $f_b = 0, 0.1, 1, 2, 4$  going from dots on the right to the left (recall that  $f_b$  parametrizes the importance of bias-squared relative to variance in the optimization). We can see that the black dots representing the linear combinations of estimators perform significantly better than single estimators; using these optimized linear combinations we are able to reach a regime with a negligible bias with respect to the noise at only a modest noise cost.



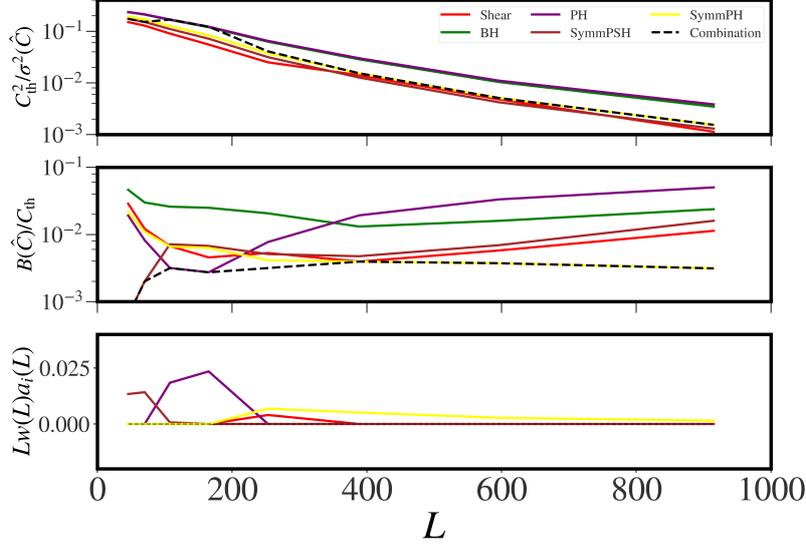
*Fig. 4.6* Composition of the optimized linear combination estimators shown in the solid black points in the previous figure. The pie charts show the fractional contribution made by each individual estimator in the linear combination, calculated as  $\int_{\mathcal{L}} w(L) a_i(L)$ . In the pie charts, colors indicate the type of estimator and shading indicates the maximum CMB multipole used in the reconstruction. When assuming  $f_b = 0$ , i.e. optimizing for minimum variance alone, as expected the standard QE alone is dominant. For an optimization which heavily penalizes bias using  $f_b = 4$ , a combination of PH and SymmPH is the best combination; this combination only has a 0.5% bias, although the signal-to-noise falls below 70. We may notice that for high  $f_b$  PH with a high  $l_{\max}$  is chosen, and not with a low one that we might expect to give lower bias. The reason is that at the important scales for the calculation of the CMB lensing amplitude, the PH with higher  $l_{\max}$  is lower in bias compared to the ones with lower  $l_{\max}$ . This is because there are internal cancellations in the calculation of the bias, depending on the specific  $l_{\max}$  (and set of simulations).



*Fig. 4.7* More detailed illustration of the estimator linear combination for the case of temperature-lensing-only power spectrum optimization with  $f_b = 1$ . The top panel shows the signal-to-noise ratio squared per mode for the lensing reconstruction, the middle panel shows the Gaussian-smoothed absolute value of the total foreground bias for each estimator. On the bottom, the weights per mode are shown for each estimator; this illustrates the relative contribution of each estimator to the linear combination. It can be seen that, for  $f_b = 1$  where bias and variance are assigned equal importance, the optimization tries to compromise between bias and noise per mode. Most of the constraining power for the CMB lensing amplitude derives from  $L \leq 500$ .

the contribution from each estimator for this case ( $f_b = 1$ ); in particular, we plot the signal-to-noise per mode, the total foreground-induced bias, and the contribution of each of the estimators in the combination. The solution is dominated by profile hardening, although it can be seen that there are contributions from SymmPSH on large scales and shear on small scales as well.

$f_b = 4$  is a case where minimizing bias takes on higher priority in the optimization. In this case, again, profile hardening is one of the dominant estimators; we obtain a solution where we mix PH at a moderate-high  $l_{\max} = 4000$  with SymmPH with CIB deprojection at low  $l_{\max} = 3000$ . Here we reduce the bias by a large factor  $\mathcal{O}(16)$ , to nearly  $\sigma/4$  or  $\approx 0.3\%$ ; however, this comes at a noise cost of around 18%, resulting in an SNR of below 70. We show the per-mode contribution from each estimator in Figure 4.8. As indicated previously, profile hardening is the dominant estimator over the most relevant range of scales; however, on small scales there is a significant contribution from SymmPH. It can be seen that the optimizer selects the estimator with the lowest bias at each scale, which is as expected since  $f_b = 4$  prioritizes bias minimization.



*Fig. 4.8* As for Figure 4.7, but here showing results for an optimization run with  $f_b = 4$ , i.e. the bias is assigned approximately four times more importance in the optimization than the noise. It can be seen that, as expected, this optimization effectively selects the estimator at each  $L$  that has the lowest bias per mode.

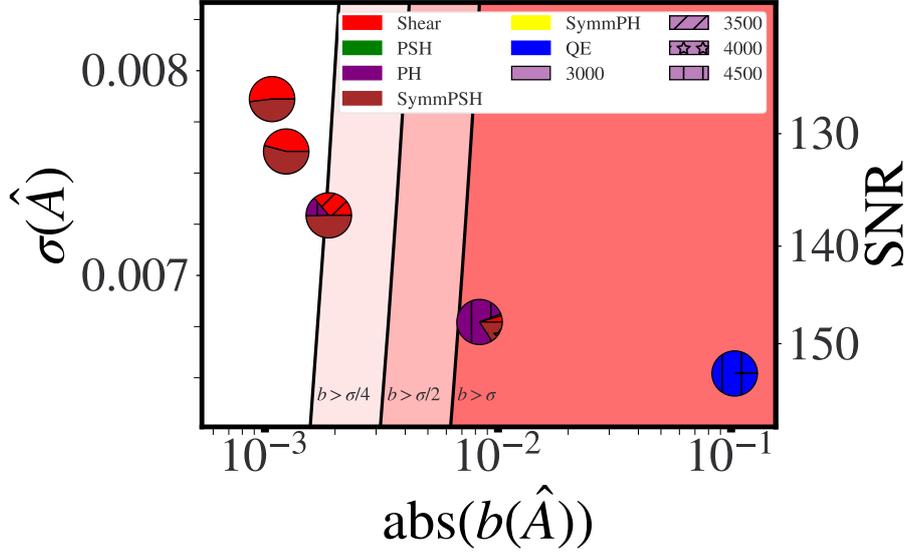
We note that for each choice of  $f_b$  we could have multiple near-optimal configurations, in the sense that some configurations with different  $l_{\max}$ s differ by a only few percent in the loss function. Since our per-mode plots are only for the single optimal configuration, we check that the three best-performing configurations are composed similarly. In Appendix (B.1) we discuss how these optimized results depend also to some extent on the masking choice.

### Minimum variance lensing reconstruction (temperature and polarization)

The situation changes significantly when including polarization data alongside temperature in the lensing estimator. Our reference case for reconstruction from both temperature and polarization will be QE plus polarization, at  $l_{TT, \max} = 3000$ ,  $l_{\text{pol}, \max} = 5000$ , for which we have a bias of 3%, with an SNR of 143. Note that in our analysis in this chapter, we assume that no foregrounds are present in the polarization maps themselves. In this case, the total foreground bias is given by

$$B = \vec{\alpha} \cdot \mathbf{B} \vec{\alpha} = \alpha_{TT}^2 B_{TT} + 2\alpha_{TT} \sum_{XY \in \text{pol}} \alpha_{XY} B^{TT, XY}, \quad (4.19)$$

where  $B_{TT}$  is the bias per mode for the  $TT$  estimator lensing power spectrum, which can in turn be written as  $\vec{a} \cdot B^{TT, TT} \vec{a}$ . Here  $B^{TT, TT}$  is the Gaussian-smoothed absolute value of the bias matrix for each temperature estimator, where we define the bias



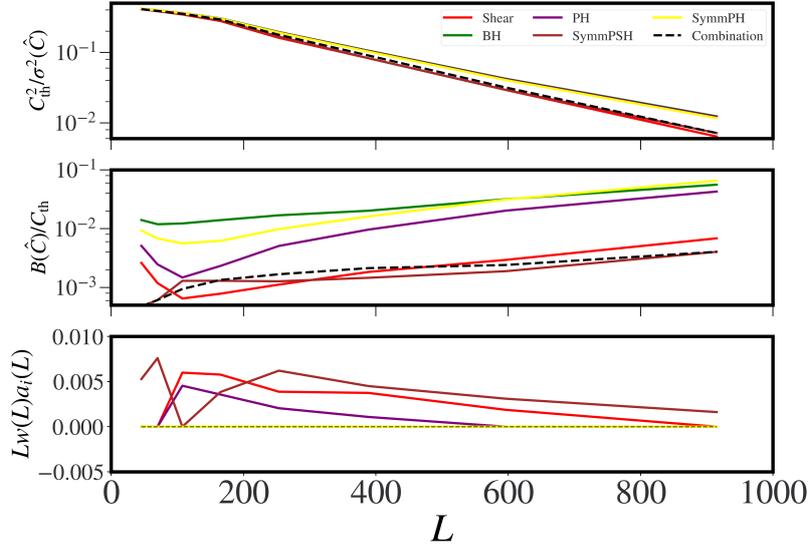
*Fig. 4.9* Linear-combination optimization results for the lensing power spectrum as in Figure 4.6; however, in this plot, we consider not just temperature data but are also including polarization data. As before, the pie charts represent the contribution from each estimator, calculated as  $\int_{\vec{L}} w(L) a_i(L)$ . We see that when bias-removal is prioritized, at left, the optimization selects a combination of geometric and multi-frequency cleaning, namely Shear and SymmPSH. This is different than the result for the optimal temperature ( $TT$ ) only combination, due to the higher importance of the primary bias term when also including polarization data as explained in the text.

matrix as a matrix with elements  $ij$  given by the bias arising from the correlation of estimator  $i$  with estimator  $j$ ; similarly,  $B^{TT,XY}$  is the smoothed absolute value of the bias matrix for each temperature estimator crossed with polarization,  $XY \neq TT$ . Finally  $\vec{\alpha}$  are minimum variance  $TT$  and polarization weights, defined in Appendix (B.1). We note that we do not allow for cancellations between the  $TT$  contribution and the polarization contribution when calculating the total bias when using both temperature and polarization data: this is done by taking the absolute value of the bias in temperature, and the absolute value of the bias with polarization data.

We find the following results for the minimum variance lensing power spectrum, as illustrated in Figure 4.9.

We immediately see a change in which estimator dominates. For higher  $f_b$ , we see that the dominant estimator is no longer PH but instead SymmPSH (with also large contributions from shear).

In particular, for  $f_b = 1$  we see that the SymmPSH-dominated,  $l_{\max} = 3000$ , combination performs very well. We find that the bias is reduced by a factor  $\mathcal{O}(17)$  with respect to QE MV with  $TT$  at  $l_{\max} = 3000$ , giving a bias that is well below percent level at around 0.18%, corresponding to just above  $\sigma/4$ . This bias reduction



*Fig. 4.10* Illustration of the estimator linear combination for power spectrum optimization with  $f_b = 1$  – now assuming both polarization and temperature data are used in the reconstruction. The top panel shows the signal-to-noise ratio squared per mode for the lensing reconstruction. The middle panel shows the Gaussian-smoothed absolute value of the total foreground bias (equation (4.19)) for each estimator and the combination. In the bottom panel, the weights per mode are shown for each estimator. It can be seen that, for  $f_b = 1$  where bias and variance are assigned equal importance, the optimization tries to compromise between bias and noise per mode.

comes at a modest noise cost of only 4%, with the total SNR still reaching nearly 137. The contributions of the different estimators are shown in Figure 4.10 for  $f_b = 1$ . We can see that SymmPSH dominates on smaller scales, although on large scales a mix of estimators, SymmPSH, SH and PH, is chosen.

For  $f_b = 4$  the bias is reduced by nearly  $\mathcal{O}(30)$  with respect to QE MV with  $TT$  at  $l_{\max} = 3000$ , giving a bias at the 0.1% level and so below  $\sigma/4$ . The noise cost of roughly 12% is still only moderate. The contributions from different estimators are shown in Figure 4.11; for  $f_b = 4$ , generally the estimator with the lowest bias, dominates, although there are some scales to which this does not apply.<sup>7</sup>

Why does the optimizer give such different results when including polarization data? The reason for this change when considering the minimum variance estimator can be understood as follows: when adding polarization data (assumed to be foreground-

<sup>7</sup>We note that the combined bias and the combined signal-over-noise squared in Figures 4.10 and 4.11 for the  $TT$ -plus-polarization case, are not just given by combining the contributions plotted for each single estimator. The reason is that, for example, the combined bias in this case is  $\vec{\alpha} \cdot \mathbf{B}\vec{\alpha}$ , as described in equation (4.19), but for each estimator, the bias plotted is  $\vec{\alpha}' \cdot \mathbf{B}'\vec{\alpha}'$ , where now we specify that the MV weights for the single estimator are different than those for the combined estimator. For this reason, in Figure 4.11 the bias is not always the smallest one among estimators.

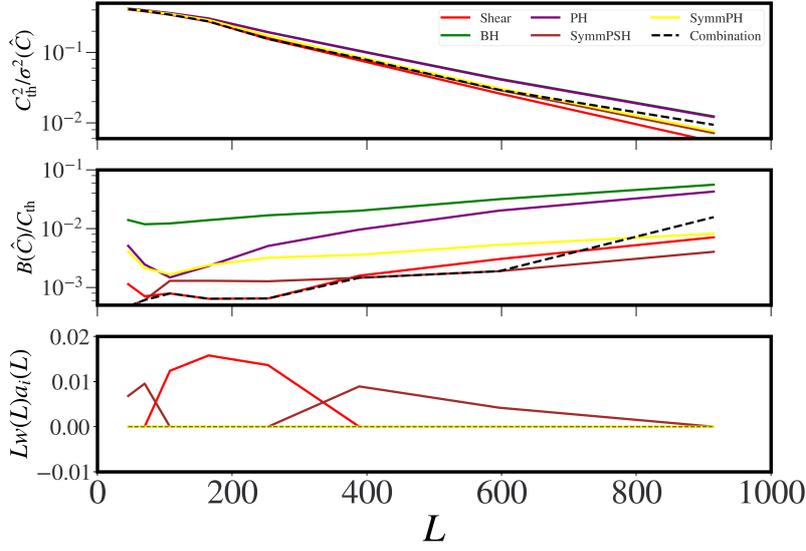


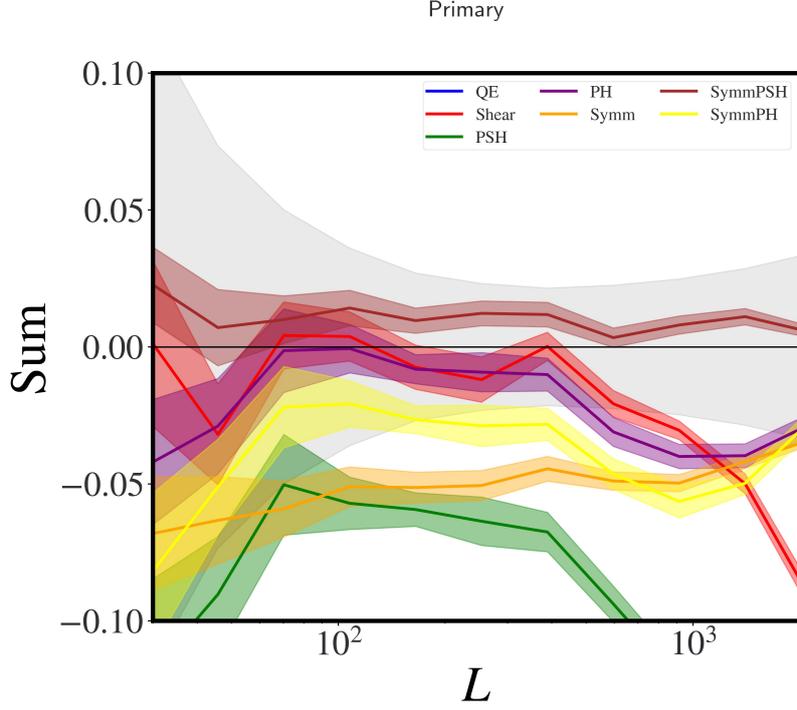
Fig. 4.11 As for Figure 4.10, but now showing results for an  $f_b = 4$  optimisation where bias reduction is prioritized. As expected, generally, the estimator with lowest bias gets selected in the combination.

free), the importance of different bias contributions from each estimator changes. As explained in Appendix (B.1), in this case the primary bias  $\langle \kappa Q_{TT+\text{pol}}[T_f, T_f] \rangle$  contribution, which arises also from the temperature-polarization estimator cross-correlation, becomes more important. This can be seen in a simple computation of the bias, assuming that polarization-only spectra do not contribute:<sup>8</sup>

$$\begin{aligned}
B &= \vec{\alpha} \cdot \mathbf{B} \cdot \vec{\alpha} \\
&= \alpha_{TT}^2 B_{TT} + 2\alpha_{TT} \sum_{XY \in \text{pol}} \alpha_{XY} B^{TT,XY} \\
&\approx \alpha_{TT}^2 B_{TT} + 2\alpha_{TT} (\alpha_{TE} B^{TT,TE} + \alpha_{EE} B^{TT,EE} \\
&\quad + \alpha_{EB} B^{TT,EB}) \quad , \quad (4.20) \\
&= \alpha_{TT}^2 B_{TT} + \alpha_{TT} (\alpha_{EE} + \alpha_{EB} + \alpha_{TE}) \frac{P_{TT}}{2} \\
&\approx \alpha_{TT}^2 B_{TT} + \frac{\alpha_{TT}(1 - \alpha_{TT})}{2} P_{TT}
\end{aligned}$$

where  $\vec{\alpha}$  are minimum variance estimator (MV) weights,  $P_{TT}$  contains only a “primary” bias contribution (arising from one of the bispectrum terms), and where we neglect the  $TB$  estimator on all scales. Since the inclusion of polarization further up-weights

<sup>8</sup>Actually, as we explain in Appendix (B.1),  $B^{TT,TE}$  does arise as a secondary contraction with the  $TE$  estimator. It is usually small on large scales, and more important on smaller scales, but where the modes are downweighted more due to the noise. For now we ignore it, although we include it in the numerical calculations. And we will completely ignore  $B^{TE,TE}$  as it enters with  $\alpha_{TE}^2$ , and  $\alpha_{TE}$  is already small.



*Fig. 4.12* The primary contribution  $\langle \kappa Q_{TT+\text{pol}}[T_f, T_f] \rangle$  relative to the CMB lensing autospectrum signal for the sum of foreground components, for  $l_{\text{max}} = 3500$  for a few estimators. This type of plot becomes important when adding polarization data, as the estimators with the lowest primary bias contribution tend to become favoured in this case, in comparison to the  $TT$ -only case.

the importance of the primary bias (which is dominant on large scales anyway), the estimators with the lowest primary bias contribution become the most important ones in the combination when polarization data is included. An example can be seen in Figure 4.12 for  $l_{\text{max}} = 3500$ ; this represents a different ordering compared to the total foreground bias shown in Figure B.13 in Appendix B.

### Cross-correlations

**Temperature-only lensing reconstruction** We now turn to optimization results for a cross-correlation with an LSST-like galaxy sample. Our reference case for  $TT$  will again be the QE at  $l_{\text{max}} = 3000$ , for which we have a bias of around 8%, with an SNR of roughly 94. Note that since the LSST galaxies do not have the same redshift distribution as the CMB lensing redshift kernel, different biases may take on a different importance; in particular, we expect that the SZ effect will become somewhat more important than CIB emission since the lower redshifts are somewhat more emphasized in this cross-correlation. We expect that this will lead to different optimization results. In Figure 4.13 we show the results for  $TT$ -only data.

For  $f_b = 1$ , the bias is reduced by a factor of 12.6 (giving a sub-percent bias of around 0.6%), at a noise cost of around 2% (still giving an SNR of more than 90), with respect to QE at  $l_{\max} = 3000$ . We see that PH dominates the combination (there is some tSZ deprojection at the CMB level through SymmPSH, although in a lower contribution with respect to PH probably due to different noise levels); in part, this may be because tSZ mitigation is of the highest importance in this cross-correlation. Going to higher  $f_b$ , PH still dominates, with some contribution from SymmPSH. For  $f_b = 4$  we are able to reduce the bias by a factor of  $\mathcal{O}(18)$  compared to QE at  $l_{\max} = 3000$ , although this comes at a price of a roughly 29% increase in noise. In the Appendix we show an example of the per mode solution for  $f_b = 1$ .

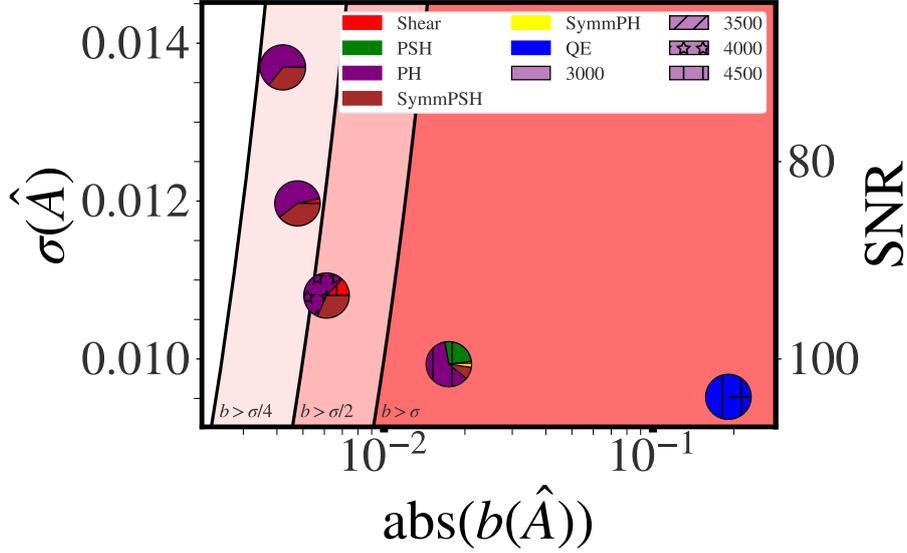
**Minimum variance estimator** Our reference case for  $TT$  plus polarization will be QE plus polarization, at  $l_{TT,\max} = 3000$ ,  $l_{\text{pol},\max} = 5000$ , for which we have a bias of 1%, with an SNR of 124.

When adding polarization data to temperature data, we see from Figure 4.14 that the optimization results are fairly similar, unlike for the autospectrum. The reason is that for cross-correlations, assuming no foregrounds are contained in the polarization maps, there is no contribution from the polarization estimators to the foreground-induced bias. However, due to the resulting lower total bias, it is possible to increase the  $l_{\max}$  of the temperature estimators, and still obtain small biases. For  $f_b = 4$ , we can see that the bias is reduced by a factor of  $\mathcal{O}(22)$  to a negligible level of less than 0.1%. This comes at only a modest, 3% noise cost compared to QE MV at  $l_{\max} = 3000$ , reaching an SNR of around 120, while remaining in a regime where the bias is negligible.

### Simplifying the estimator combinations

A question that arises is to what extent the complex estimator combinations considered previously can be simplified without degrading their performance. Indeed, some of the optimal estimator combinations we derived above may be burdensome to implement in practice, as they require implementing multiple lensing quadratic estimators, applied to temperature maps from several multi-frequency linear combinations. We therefore investigate how much these complex estimators improve over simple two-estimator combinations.

In particular, we omit all but the two most-contributing estimators in an optimized linear combination, and we recalculate the weights for the two most-contributing estimators by maintaining their relative proportion in the  $a(\vec{L})$  weight, and then recalculating  $w(\vec{L})$  (note this is not equivalent to an optimization over two estimators).



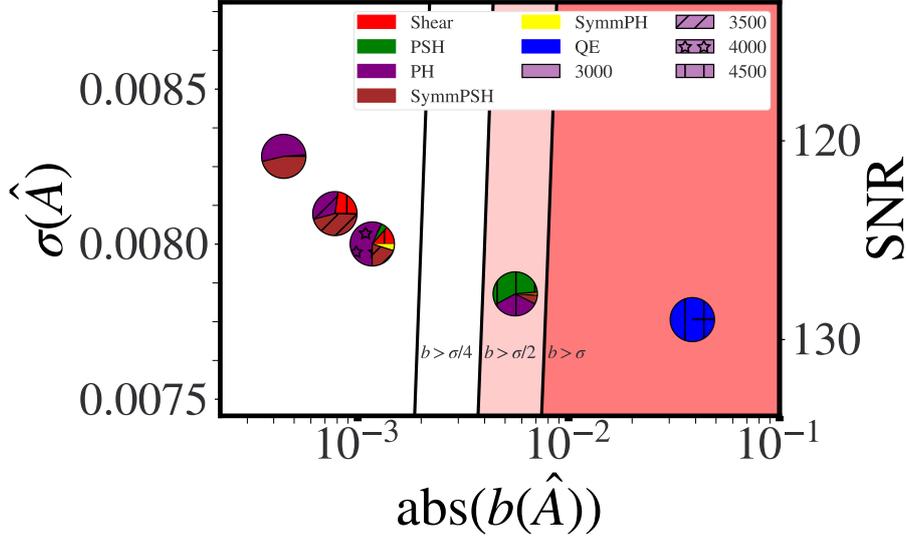
*Fig. 4.13* Optimization results for the cross-correlation of lensing from temperature-only ( $TT$ ) data with an LSST-like sample. The pie charts represent the contribution from each estimator, calculated as  $\int_{\mathcal{L}} w(L) a_i(L)$ . We see that for a high deprojection of the bias, at far left, the optimization selects a combination of geometric and multi-frequency cleaning, namely PH and SymmPSH.

The results are summarized in Figure 4.15 for the auto-spectrum, and Figure 4.16 for the cross-spectrum. We find that in most cases, this simplification has very little impact on the estimator performance: we typically find only small, percent-level shifts in bias and noise on the amplitude (we also plot the performance of some individual estimators for comparison).

For high- $f_b$  estimators, we also find that as a general rule of thumb, these are composed from the two single estimator configurations with lowest bias. In particular, SymmPSH and PH are favoured for cross-correlations with low- $z$  tracers, which can be understood as follows: PH is ideal for eliminating tSZ effects and SymmPSH, with tSZ deprojected, performs well at eliminating both tSZ and CIB biases; both together are sufficient to reach a regime of  $b < \sigma/2$ . For the auto-spectrum these two estimators similarly form a robust combination.

## 4.6 Conclusions

In this chapter we have investigated which combination of different lensing foreground mitigation methods minimizes the foreground biases while maximizing the signal-to-noise ratio of CMB lensing power spectrum and cross-spectrum measurements.



*Fig. 4.14* Optimization results for the cross correlation of lensing from polarization and temperature data with an LSST-like sample. The pie charts represent the contribution from each estimator, calculated as  $\int_{\mathcal{L}} w(L) a_i(L)$ . As the primary cross-bias has no contributions from (foreground-free) polarization, we can see that for high deprojection of the bias, at far left, the results are very similar to the  $TT$  only cross-correlation case.

We explore two different approaches to combine mitigation methods for this purpose.

The first one is composing (i.e., simultaneously using) geometric and multi-frequency deprojection methods, namely bias hardening applied to a symmetric estimator in which a particular foreground has been removed via multifrequency deprojection. In this way, we propose two new estimators: i) SymmPSH, given by point source hardening applied to a symmetric estimator with tSZ deprojected; ii) SymmPH, given by profile hardening applied to a symmetric estimator with CIB deprojected.

The second approach, used to further mitigate foregrounds, is to linearly combine different types of foreground-reducing lensing estimators, and minimize a loss function given by the sum of the squares of the noise and bias on the lensing amplitude. We find that when using both temperature and polarization data, assuming no intrinsic foregrounds in the polarization, the optimal combination for a high reduction in bias is given by a mixture containing mainly point source hardening applied to the tSZ-deprojecting symmetric estimator; this gives a total bias that is below a half-percent (more than an order of magnitude smaller than the QE bias), at only a modest 11% increase in noise beyond the QE. For cross-correlating with an LSST-like galaxy sample, we find that the best combination for a high reduction in bias is given

by a mixture of point source hardening applied to the CIB-deprojecting symmetric estimator combined with the profile hardened estimator on its own; this results in a bias to the cross-correlation amplitude that is less than 0.1% (significantly smaller than the QE bias) at a cost of only a 3% increase in noise.

While the exact combination depends on the observable considered, generally we find that the most robust estimators that can be used for auto and cross-spectrum analyses consist mainly of combinations of: the profile hardened estimator, point source hardening applied to the tSZ-deprojecting symmetric estimator, and, although it has higher noise, the shear estimator (SH).

We caution that the results of this chapter may have some dependence on the specific set of simulations used, although we have made efforts to perform our analysis without sensitivity to precise cancellations or other fine details. In this chapter, we have focused on the simulations of [244], but exploring our results with other sets of simulations, such as as Websky [268], is well motivated. Future work could also include both extragalactic and galactic foregrounds in the polarization map simulations or – departing from simulations altogether – a theoretical modeling of foreground biases. We defer such investigations to future work.

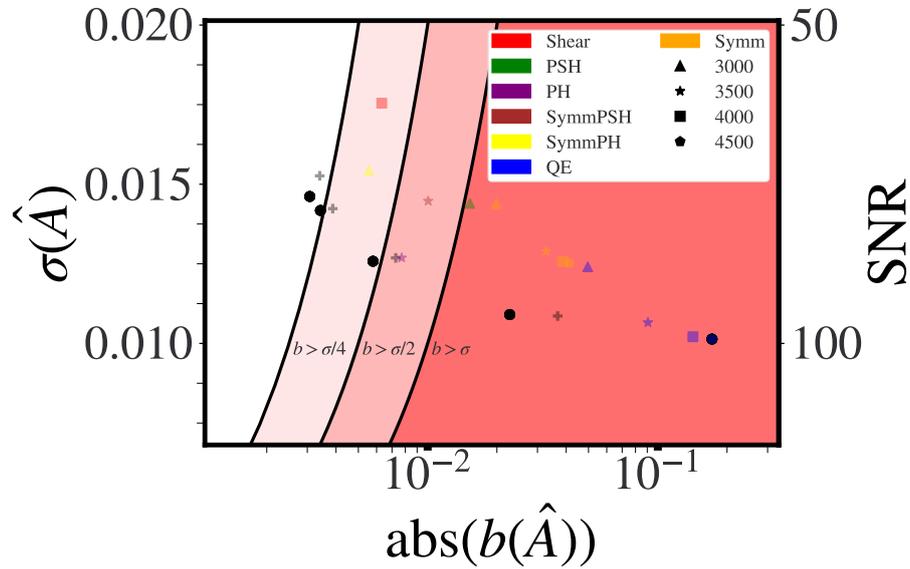
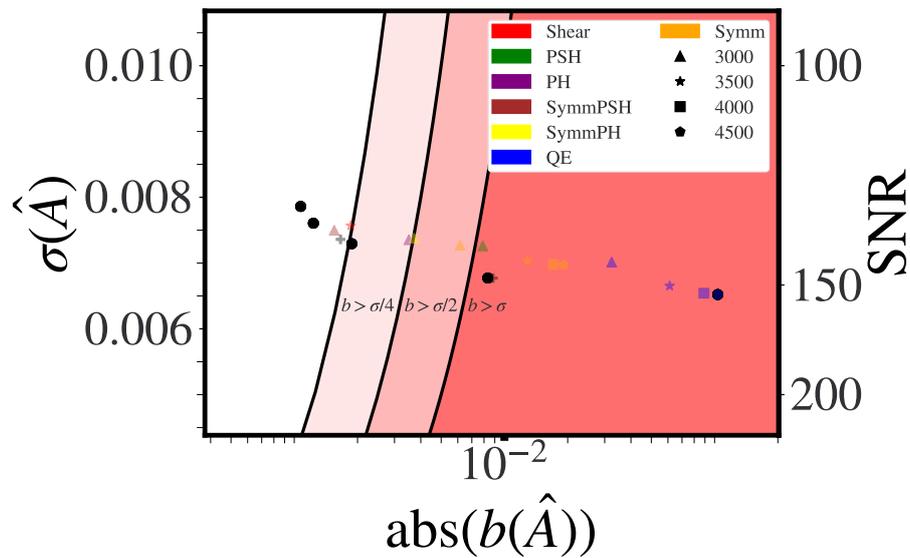
(a)  $TT$ -only case.(b)  $TT$ +polarization case.

Fig. 4.15 Can the (sometimes complex) optimal linear combinations of estimators be easily simplified? Black dots show the optimal points for the CMB lensing auto-spectrum shown in previous plots. The grey cross is the shift in this optimal point, if we choose to simplify the combination by using only the two estimators which contribute most. In the top panel we show the  $TT$ -only case, and in the bottom panel we also include polarization data. We can see that in general we have at most percent level shifts, going from black dot to grey cross, so that simplifying the estimator combinations works well.

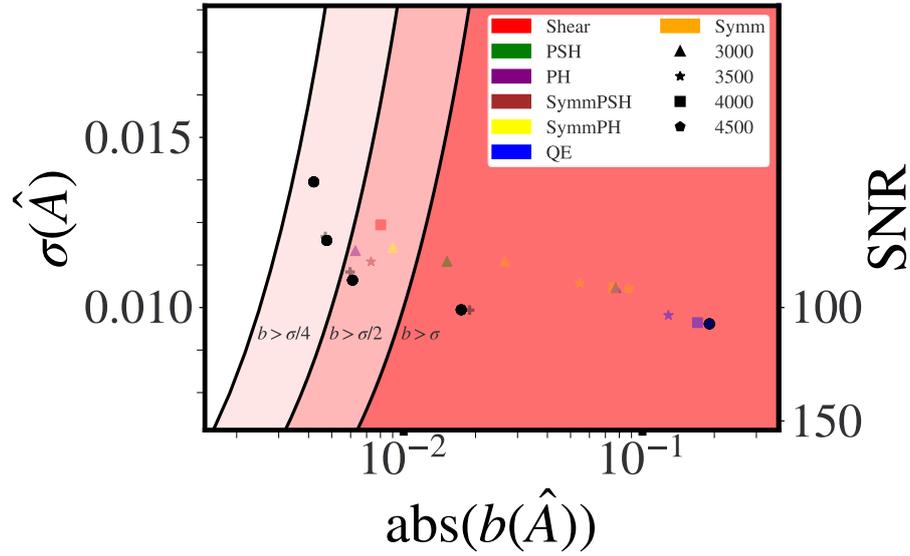
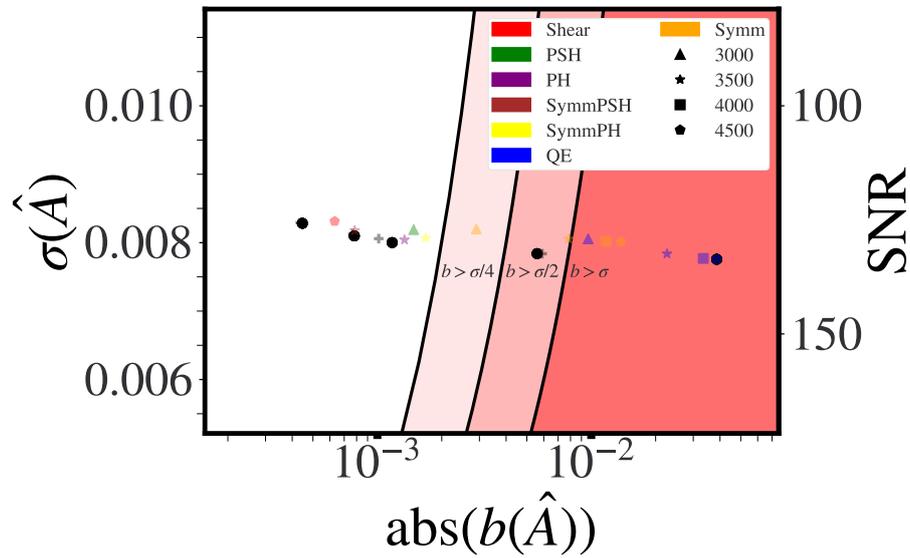
(a) *TT*-only case.(b) *TT*+polarization case.

Fig. 4.16 As for Figure 4.15, but for cross-correlations with LSST galaxies. We again conclude that the combined estimators can be simplified without significantly degrading their performance.

## Density reconstruction from biased tracers and its application to primordial non-Gaussianity

### Summary

*Large-scale Fourier modes of the cosmic density field are of great value for learning about cosmology because of their well-understood relationship to fluctuations in the early universe. However, cosmic variance generally limits the statistical precision that can be achieved when constraining model parameters using these modes as measured in galaxy surveys, and moreover, these modes are sometimes inaccessible due to observational systematics or foregrounds. For some applications, both limitations can be circumvented by reconstructing large-scale modes using the correlations they induce between smaller-scale modes of an observed tracer (such as galaxy positions). In this chapter, we further develop a formalism for this reconstruction, using a quadratic estimator similar to the one used for lensing of the cosmic microwave background. We incorporate nonlinearities from gravity, nonlinear biasing, and local-type primordial non-Gaussianity, and verify that the estimator gives the expected results when applied to  $N$ -body simulations. We then carry out forecasts for several upcoming surveys, demonstrating that, when reconstructed modes are included alongside directly-observed tracer density modes, constraints on local primordial non-Gaussianity are generically tightened by tens of percents compared to standard single-tracer analyses. In certain cases, these improvements arise from cosmic variance cancellation, with reconstructed modes taking the place of modes of a separate tracer, thus enabling an effective "multitracer" approach with single-tracer observations.*

## 5.1 Introduction

Our understanding of the Universe has benefited tremendously from measurements of the cosmic microwave background (CMB), primarily because of the linear relationship between fluctuations in the CMB and fluctuations generated in the very early universe. This relationship allows us to connect CMB measurements to the statistics of the initial fluctuations and their time evolution, and has led to the establishment of the current cosmological model. Extraction of similar information from the large-scale structure (LSS) of the universe is limited by nonlinear clustering at smaller distances and lower redshifts, requiring more elaborate modelling to interpret observations. This modelling burden is greatly reduced at the largest distances we can resolve with galaxy surveys, but this regime is in turn obscured by both statistical and systematic errors. In this chapter, we explore a method to access these large scales while bypassing both types of errors: quadratic density reconstruction.

This idea of density reconstruction relies on the fact that a fixed long-wavelength density fluctuation correlates two different small-scale modes due to non-linear evolution and higher-order biasing, with the amount of correlation proportional to the long-wavelength mode. This can be understood as arising from a violation of statistical homogeneity if the long-wavelength mode is considered as fixed and the shorter-wavelength modes are averaged over in an ensemble. Writing down a quadratic estimator that probes this induced correlation between two different modes, we can estimate the long-wavelength modes from the statistical properties of the smaller-scale modes.<sup>1</sup>

There is a close analogy between this procedure and the common method of CMB lensing reconstruction, in which a quadratic estimator, making use of the lensing-induced correlation between two different CMB temperature modes, is used to reconstruct the lensing field (e.g. [128]). It is using this analogy that many of the methods for density reconstruction were derived. The idea of using a standard quadratic estimator in the CMB lensing form to perform this reconstruction was first proposed by [91], building on earlier work ([206, 292, 293], albeit with a somewhat modified estimator). Significant further work in this area has been presented by [155, 188, 141, 157, 156]; see further discussion in Section 5.5.

---

<sup>1</sup>In fact, these statements are independent of the relative wavelengths of the modes, and the formalism we present in this chapter is not restricted to the so-called “squeezed limit” of the three modes involved. However, for our applications, the modes we are seeking to reconstruct have longer wavelengths than the two modes whose correlations are used for the reconstruction, so we focus on that situation in this chapter.

The work in this chapter broadly divides in two parts. In the first part, we present a considerable expansion of current technology for density reconstruction. We discuss the application of density reconstruction to biased tracers, including, for the first time, a full non-linear bias model in such a formalism. We further validate our method on a suite of realistic  $N$ -body simulations, demonstrating that our methods perform just as expected from theoretical calculations for both the reconstruction and its noise level.

We note that this reconstruction has a wealth of applications. One simple application is the following: LSS surveys are often plagued by observational systematics that manifest at large scales, impeding the direct observation of low- $k$  modes. Galaxy and quasar surveys are affected, for example, by variations in the density of foreground stars, seeing, and galactic dust extinction (e.g. [225, 119, 139]), while 21 cm surveys cannot access modes with low line-of-sight wavenumbers that are dominated by galactic foregrounds, and imperfect knowledge of the instrument can spread this contamination throughout a wider region of Fourier space (e.g. [204, 161, 162]). A method of reconstructing these inaccessible modes using correlations between smaller-scale modes will improve the constraining power of a given survey for large-scale signals such as local non-Gaussianity, and allow cross-correlations involving 21 cm surveys that would otherwise be impossible (e.g. [155]). In this chapter, we parameterize large-scale systematics with a wavenumber  $K_{\min}$  below which the tracer modes are assumed to be inaccessible, and explore the precision with which modes with  $K < K_{\min}$  can be recovered by our estimator. We note that this assumes that the relevant systematics can be parameterized as a large-scale additive component, rather than a possible modulation that might also significantly affect small scales; while there is evidence that this is a reasonable assumption for some of the currently known systematics (e.g. [139]), it may not hold in all cases.

In contrast to this general application, the second goal of our chapter is to explore, in detail, a much more subtle application of density reconstruction: improving constraints on local-type primordial non-Gaussianity. We will briefly motivate the measurement of primordial non-Gaussianity and the utility of density reconstruction for improving these constraints in the following paragraphs.

The CMB has taught us that the two-point statistics of the primordial fluctuations can be accurately described by a red-tilted power law in Fourier space. If the initial conditions are completely described by this power law, they have to be Gaussian distributed, with statistics determined by only two degrees of freedom: the amplitude ( $A_s$ ) and tilt ( $n_s$ ) of the power law. If this is the case, however, it will be difficult to reach beyond our current understanding of the early Universe. The most widely accepted theory is known as cosmic inflation, which postulates a short early period of accelerated cosmic expansion. Effectively, the constraints we derive from the

CMB tell us that inflation can be very well described by a scalar field slowly rolling down a potential (“single-field slow roll”, or SFSR), with only (weak) gravitational interactions. While such a model is certainly possible (it was the first to be considered [104, 159, 160]), it will not provide us with simple opportunities to understand the physics of inflation. If a proposed model of the early Universe has to comply with Gaussian initial conditions, effectively the model will observationally resemble SFSR. Any further distinction could be extracted from the details of the scale dependence of the primordial power spectrum [259], but so far, observations do not reveal any obvious deviations from a single-parameter power-law [12, 11].

A much more powerful model discriminator would be available if the initial conditions showed a (small) deviation from Gaussianity. In the presence of non-Gaussianity, all moments beyond the power spectrum will generically be excited (starting with the 3-point function or bispectrum). Technically, these higher-point spectra probe the dynamics and interactions of the field(s) driving inflation. As a result, a measurement of non-Gaussianity would reveal details of inflation that can be directly related to the underlying fundamental physics. For example, non-Gaussianity could reveal the presence of more fields relevant during inflation, or could provide clues to how strongly coupled the inflation field is (see e.g. [183] and references therein). These powerful constraints cannot be exposed through any other measurement, making non-Gaussianities a unique probe of the early Universe.

To lowest order, primordial non-Gaussianities modulate the gravitational potential  $\Phi$  via

$$\Phi(\vec{k}) = \varphi_G(\vec{k}) + f_{\text{NL}}^X \int \frac{d^3q}{(2\pi)^3} G_{\text{NL}}^X(\vec{q}, \vec{k} - \vec{q}) \varphi_G(\vec{q}) \varphi_G(\vec{k} - \vec{q}), \quad (5.1)$$

where  $\varphi_G$  is the Gaussian potential and  $G_{\text{NL}}^X$  is a kernel that describes how the potential is modulated. In this chapter, we are interested in local non-Gaussianities for which  $G_{\text{NL}}^{\text{local}} = 1$ , i.e.

$$\Phi(\vec{x}) = \varphi_G(\vec{x}) + f_{\text{NL}}(\varphi_G^2(\vec{x}) - \langle \varphi_G^2 \rangle), \quad (5.2)$$

where we have subtracted the mean to yield zero expectation value for the fluctuations and have renamed  $f_{\text{NL}}^{\text{local}}$  to  $f_{\text{NL}}$ . Current constraints set  $\sigma(f_{\text{NL}}) \sim \mathcal{O}(5)$  [11], while  $f_{\text{NL}} \sim 1$  has been identified as a compelling theoretical threshold [17] which provides a strong motivation to go beyond current limits: if a measurement is made showing  $f_{\text{NL}}$  above this limit, it would effectively rule out SFSR inflation as a viable scenario. Future ground-based CMB experiments [5, 64] may be able reach  $\sigma(f_{\text{NL}}) \sim \mathcal{O}(1)$ , but poor scaling and galactic and cosmological foregrounds will likely prevent the CMB from reaching (far) beyond this limit. Fortunately, the large scale structure (LSS) in the universe provides access to many more modes, for which  $\sigma(f_{\text{NL}}) \propto$

$(k_{\max}^3 \log k_{\max}/k_{\min})^{-1/2}$  [240]. While increased dimensionality will help to improve constraints, the use of LSS will introduce many complications. For one, the scaling argument breaks down when  $k_{\max}$  exceeds the nonlinear scale  $k_{\text{NL}}$ , which is of order  $0.2h \text{ Mpc}^{-1}$  for current galaxy surveys [70, 134]. Furthermore, line-of-sight information, which will be crucial in obtaining a sufficient number of modes, will require a careful treatment mainly due to redshift-space effects [99]. Obtaining cosmological constraints from a measurement of the full LSS bispectrum will therefore be challenging, not least because of non-Gaussian covariance [240, 243, 142] which will likely require (a large number of) simulations to estimate [55]. Some of these difficulties can be overcome by simplifying the full bispectrum into more compressed statistics [236, 88, 46, 68, 189, 60, 218, 102]. The advantage of these statistics is that they should capture nearly all the information [189], but are computationally and observationally less challenging.

Unlike in the CMB, in LSS local primordial non-Gaussianity can also significantly affect the power spectrum of biased tracers, such as galaxies. Specifically, it has been shown [69, 176, 260, 78, 234] that tracer bias will be affected by the primordial non-Gaussianity, with the bias acquiring a unique  $1/k^2$  contribution, which is hard to produce otherwise. This signature has been used to place constraints on  $f_{\text{NL}}$  with current surveys [98, 150, 52]. Unfortunately, although the signal should be distinguishable from other effects, on large scales, the precision with which we can measure the power spectrum is ultimately limited by cosmic variance from the number of available modes. However, it was shown that this cosmic variance can be mitigated [246, 179, 105, 237, 165] by using multiple tracers of the same underlying density field (with different biases), which essentially allows a measurement of scale-dependent bias via a mode-by-mode comparison of the different tracers. A combination of two (or more, e.g. [237, 26]) tracers will allow for cosmic variance cancellation, limiting a measurement of the scale-dependent bias from local primordial non-Gaussianity only by the number density of these tracers. Forecasts show that these techniques enable constraints to reach  $\sigma(f_{\text{NL}}) \sim 1$  this decade [237, 26, 192, 5].

In this chapter, we show that this cosmic variance cancellation can also be achieved, to some extent, using only a single tracer. In order to do this, we compare our reconstructed density field (which provides information from higher-point functions) with a directly-measured tracer field. In the end, the constraints on  $f_{\text{NL}}$  will depend on the auto-correlation of the tracer field  $P_{\text{gg}}$ <sup>2</sup>, the cross correlation of the tracer and the reconstructed field  $P_{\text{gr}}$ , and the auto-correlation of the reconstructed field  $P_{\text{rr}}$ . This idea is related to [218], where similar ideas are used to simplify a forecast

<sup>2</sup>Since we will focus on the use of galaxies as tracers in this chapter, we will use the subscript g to refer to these tracers, although the method we describe is equally applicable to quasars, line intensity maps, or other tracers.

of the combined information in the power spectrum, bispectrum and trispectrum. However, unlike in [218], we examine the reconstruction approach as a possible analysis tool rather than a method for more easily computing complex forecasts. In addition, whereas [218] relies on an extension of position-dependent power spectra [61, 59, 58, 6], which draw information only from the squeezed limit, here we use a quadratic estimator formalism for the reconstructed field without imposing a squeezed-limit constraint.

Let us briefly summarize our most important results:

- The modes of the tracer overdensity will be coupled due to nonlinearities from gravity, nonlinear bias, and primordial non-Gaussianity. The amplitudes (parameterized with bias coefficients) of several of these mode-couplings are unknown a priori. We incorporate this in our characterization of the quadratic estimator for long-wavelength modes, and marginalize over the unknown coefficients in our forecasts. We also highlight the important contribution of tracer shot noise to the noise on the reconstructed modes.
- We demonstrate density reconstruction using dark matter halos in  $N$ -body simulations, verifying that the performance agrees well with that predicted from analytical formulas. Though additional work using simulations will be required for a practical analysis, our results indicate that our forecasts are realistic.
- We show that the quadratic estimator is able to reconstruct long-wavelength modes at high signal-to-noise for a wide range of upcoming surveys (see Fig. 5.8).
- The addition of the reconstructed field to forecasts using the large-scale biased tracer field can improve constraints on  $f_{\text{NL}}$  by tens of percents depending on the survey configuration. The improvement arises from a combination of two sources: sample variance cancellation of signal in the large-scale tracer field, and additional scale-dependent signal in the reconstructed field on scales where the tracer field may be obscured by observational systematics. The additional information in the reconstructed modes can be viewed as a signature of non-Gaussian signal in the three and four-point functions, and our approach can be viewed as a simple method to obtain combined information from the three- and four-point functions and the power spectrum.
- The performance of this approach to constraining  $f_{\text{NL}}$  is limited by a combination of tracer number density and maximum wavenumber of modes that can be used for reconstruction, with the details again depending on the survey configuration. Potential improvements using response function approaches [29, 28] could be explored to extend the reconstruction wavenumber and gain signal-to-noise.

Quantity	Symbol	Defined in
Dirac delta function in 3d	$\delta_{\text{D}}(\vec{k})$	—
Wavenumbers of modes used in reconstruction	$\vec{k}, \vec{q}, \text{etc.}$	—
Wavenumbers of modes used for $f_{\text{NL}}$ constraints	$\vec{K}, \vec{K}', \text{etc.}$	—
Amplitude of local primordial non-Gaussianity	$f_{\text{NL}}$	Eq. (5.2)
Factor relating primordial potential and $\delta_1$	$M(k, z)$	Eqs. (5.19)-(5.20)
Linear matter overdensity	$\delta_1(\vec{k}, z)$	—
Linear matter power spectrum	$P_{\text{lin}}(k, z)$	—
Tracer overdensity	$\delta_{\text{g}}(\vec{k}, z)$	Eq. (5.3) [generic]; Eq. (5.37) [second-order bias model]
Second-order mode-coupling	$F_{\alpha}(\vec{k}_1, \vec{k}_2)$	Eq. (5.3) [generic]; Eq. (5.37) [second-order bias model]
Second-order response of small-scale power spectrum to long mode	$f_{\alpha}(\vec{k}_1, \vec{k}_2, z)$	Eq. (5.10)
Coefficient of $F_{\alpha}$ in second-order bias model for $\delta_{\text{g}}$	$c_{\alpha}$	Eq. (5.37)
Linear tracer bias	$b_1 \equiv b_{10}^{\text{E}}$	Eq. (5.37)
Quadratic tracer bias	$b_2 \equiv b_{20}^{\text{E}}$	Eq. (5.37)
Other second-order bias parameters	$b_{s2}^{\text{E}}, b_{01}^{\text{E}}, \dots$	Sec. 5.2.2
Quadratic estimator for mode with wavenumber $\vec{K}$	$\hat{\Delta}_{\alpha}(\vec{K})$	Eqs. (5.11), (5.17)
Weight function in $\hat{\Delta}_{\alpha}(\vec{K})$	$g_{\alpha}(\vec{k}_1, \vec{k}_2)$	Eq. (5.15)
Normalization and Gaussian noise of $\hat{\Delta}_{\alpha}(\vec{K})$	$N_{\alpha\beta}(\vec{K})$	Eq. (5.16)
Mode reconstructed with growth-coupling estimator $\hat{\Delta}_{\text{G}}(\vec{K}, z)$	$\hat{\delta}_{\text{r}}(\vec{k}, z)$	—
Power spectrum of $\delta_{\text{g}}$ , ignoring shot noise contribution	$P_{\text{gg}}$	—
Sum of $P_{\text{gg}}$ and shot noise contribution	$P_{\text{tot}}$	Sec. 5.2.1
Cross power spectrum between $\delta_{\text{g}}$ and $\hat{\delta}_{\text{r}}$ , ignoring shot noise contribution	$P_{\text{gr}}$	—
Power spectrum of $\hat{\delta}_{\text{r}}$ , ignoring shot noise contribution	$P_{\text{rr}}$	—
Shot noise contribution to $\delta_{\text{g}}$ power spectrum	$P_{\text{gg,shot}}$	Eq. (5.50)
Shot noise contribution to $\hat{\delta}_{\text{r}}$ power spectrum	$P_{\text{rr,shot}}$	Eqs. (C.26)-(C.27)
Shot noise contribution to $\delta_{\text{g}}\text{-}\hat{\delta}_{\text{r}}$ cross power spectrum	$P_{\text{gr,shot}}$	Eqs. (C.31)-(C.32)
Lowest wavenumber within survey volume	$K_{\text{f}}$	Sec. 5.4.3
Wavenumber below which we assume $\delta_{\text{g}}$ cannot be measured	$K_{\text{min}}$	Sec. 5.4.3
Maximum wavenumber used for $f_{\text{NL}}$ constraints	$K_{\text{max}}$	Sec. 5.4.3
Maximum wavenumber used in quadratic estimator for reconstructed modes	$k_{\text{max}}$	Sec. 5.4.3

Table 5.1 Notation used for important quantities in this chapter

The outline of the chapter is as follows. In Section 5.2, we describe our methodology for density reconstruction, including the quadratic estimator formalism and bias expansion we use. In Section 5.3, we apply this method to halos in  $N$ -body simulations. In Section 5.4, we present our forecasts for the expected precision on reconstructed modes, as well as constraints on local non-Gaussianity. We compare this reconstruction formalism to other work involving higher-point statistics in Section 5.5. Finally, we conclude in Section 5.6. Several derivations and technical details are included in the appendices, and a summary of our notation can be found in Table 5.1. Except for in Sec. 5.3, we use cosmological parameters from the Planck 2015 results, given in the ‘‘TT,TE,EE+lowP+lensing+ext’’ column of Table 4 of [4].

## 5.2 Density reconstruction

### 5.2.1 Quadratic estimator: general case

In this section, we will develop the general formalism for reconstructing large-scale<sup>3</sup> density modes using observations of a biased tracer. This is largely based on the treatment in [91], but we have adapted their expressions to 3D wavenumbers rather than a separate treatment of line-of-sight and transverse components of  $\vec{k}$ .

Suppose that the overdensity field of the tracer,  $\delta_{\mathbf{g}}$ , is well-described by a linear bias with respect to the linear matter overdensity  $\delta_1$ , plus a set of quadratic terms that couple modes of  $\delta_1$  with kernels  $F_\alpha$  and amplitudes  $c_\alpha$ :

$$\delta_{\mathbf{g}}(\vec{k}, z) \approx b_1(z)\delta_1(\vec{k}, z) + \sum_{\alpha} c_{\alpha}(z) \int_{\vec{q}} F_{\alpha}(\vec{q}, \vec{k} - \vec{q}; z) \delta_1(\vec{q}, z) \delta_1(\vec{k} - \vec{q}, z), \quad (5.3)$$

where

$$\int_{\vec{q}} \equiv (2\pi)^{-3} \int d^3\vec{q}. \quad (5.4)$$

For example, if we took  $\delta_{\mathbf{g}}$  to be the matter overdensity rather than a biased tracer, we would have  $b_1 = 1$  and the sum would run over the second-order mode-couplings induced by gravitational evolution, which take the form (e.g. [252])

$$F_{\mathbf{G}}(\vec{k}_1, \vec{k}_2; z) \equiv \frac{17}{21}, \quad F_{\mathbf{S}}(\vec{k}_1, \vec{k}_2; z) \equiv \frac{1}{2} \left( \frac{1}{k_1^2} + \frac{1}{k_2^2} \right) \vec{k}_1 \cdot \vec{k}_2, \\ F_{\mathbf{T}}(\vec{k}_1, \vec{k}_2; z) \equiv \frac{2}{7} \left[ \frac{(\vec{k}_1 \cdot \vec{k}_2)^2}{k_1^2 k_2^2} - \frac{1}{3} \right], \quad (5.5)$$

with  $c_{\mathbf{G}} = c_{\mathbf{S}} = c_{\mathbf{T}} = 1$  and the subscripts indicating that these functions arise from isotropic **G**rowth, a large-scale coordinate **S**hift, and a **T**idal coupling. For a biased tracer, nonlinear biasing will lead to  $c_\alpha \neq 1$  for the above couplings, and primordial non-Gaussianity will introduce additional mode-couplings. In Sec. 5.2.2, we will introduce the full set of mode-couplings that must be considered, but we note here that many of the corresponding  $c_\alpha$  coefficients will not be known a priori, and this must be accounted for in the density reconstruction procedure. Henceforth, we will drop the  $z$ -dependence from the quantities defined above.

Now, we would like to use the mode-couplings in Eq. (5.3) to construct a quadratic estimator for a given mode of  $\delta_1$ . We will present the logic in some detail, for readers

<sup>3</sup>We again remind the reader that the our formalism is generally applicable, without any strong assumptions on the wavelengths of the modes.

who may not be familiar with the relevant arguments, but a reader who is comfortable with peak-background-split arguments or the CMB lensing formalism may wish to skip to the final result in Eqs. (5.9)-(5.10).

The analogous procedure for CMB lensing is to first consider an ensemble average over CMB fluctuations while keeping fluctuations in the lower-redshift matter density fixed. In this case, the fixed modes of the lensing potential  $\phi$  (which is a line-of-sight projection of the lower-redshift density field – see e.g. [128]) break the statistical isotropy of the CMB fluctuations, inducing correlations between CMB fluctuation modes with different wavenumbers: for temperature modes on the flat sky, the specific effect is given by

$$\langle T(\vec{\ell})T(\vec{L} - \vec{\ell}) \rangle_{\phi \text{ fixed}} = (2\pi)^2 \delta_{\text{D}}(\vec{L}) C_L + f_{\phi}(\vec{\ell}, \vec{L} - \vec{\ell}) \phi(\vec{L}) . \quad (5.6)$$

When analyzing CMB simulations or data, the temperature two-point function is estimated by a (weighted) sum over  $\vec{\ell}$  within a given CMB realization, and this in fact approximates the ensemble average above, with  $\phi$  modes effectively fixed because they do not explicitly enter the sum. Equation (5.6) is an efficient starting point for deriving quadratic estimators for a specific mode of  $\phi$ , and we would like to find the analogous starting point for density reconstruction.

To proceed, we consider an ensemble average over all modes of  $\delta_1$  except those with wavenumbers in a small neighborhood around  $\vec{K}$ , with  $\delta_1(\vec{K})$  being the mode we will eventually want to reconstruct. (We must consider a neighborhood around  $\vec{K}$  because we are working in the continuum limit, where we have integrals instead of discrete sums over wavenumbers; we will return to this point below.) In this ensemble average, which we denote by “ $\sim \vec{K}$  fixed”, and using Eq. (5.3), the two-point function of  $\delta_g$  is at next-to-leading-order in  $\delta_1$  is

$$\begin{aligned} \langle \delta_g(\vec{k}) \delta_g(\vec{K} - \vec{k}) \rangle_{\sim \vec{K} \text{ fixed}} &= b_1^2 \langle \delta_1(\vec{k}) \delta_1(\vec{K} - \vec{k}) \rangle \\ &+ b_1 \int_{\vec{q}} \sum_{\alpha} c_{\alpha} F_{\alpha}(\vec{q}, \vec{k} - \vec{q}) \langle \delta_1(\vec{q}) \delta_1(\vec{k} - \vec{q}) \delta_1(\vec{K} - \vec{k}) \rangle_{\sim \vec{K} \text{ fixed}} \\ &+ [\vec{k} \leftrightarrow \vec{K} - \vec{k}] . \end{aligned} \quad (5.7)$$

In the first line, we have assumed that  $\vec{k}$  is not within the chosen neighborhood of  $\vec{K}$  or the equivalent neighborhood of 0, so there is no difference between our special ensemble average and the standard one. In the second line, the integrand evaluates to zero if  $\vec{q}$  and  $\vec{k} - \vec{q}$  are not within the neighborhood of  $\vec{K}$ , since in that case, all three  $\delta_1$  modes are averaged over, and the three-point function is zero for  $\vec{K} \neq 0$ . When  $\vec{q} \sim \vec{K}$  or  $\vec{k} - \vec{q} \sim \vec{K}$ , where we use “ $\sim \vec{K}$ ” to indicate a vector falling within the

neighborhood of  $\vec{K}$ , then  $\delta_1(\vec{q})$  or  $\delta_1(\vec{k} - \vec{q})$  factor out of the ensemble average because they are held fixed, and the remaining two modes are averaged over:

$$\begin{aligned} \langle \delta_g(\vec{k}) \delta_g(\vec{K} - \vec{k}) \rangle_{\sim \vec{K} \text{ fixed}} &= b_1^2 \langle \delta_1(\vec{k}) \delta_1(\vec{K} - \vec{k}) \rangle \\ &+ 2b_1 \int_{\vec{q} \sim \vec{K}} \sum_{\alpha} c_{\alpha} F_{\alpha}(\vec{q}, \vec{k} - \vec{q}) \delta_1(\vec{q}) \langle \delta_1(\vec{k} - \vec{q}) \delta_1(\vec{K} - \vec{k}) \rangle + \\ &+ [\vec{k} \leftrightarrow \vec{K} - \vec{k}]. \end{aligned} \quad (5.8)$$

From here, we simply evaluate the two-point correlators and use the resulting Dirac delta functions to collapse the  $\vec{q}$  integrals.<sup>4</sup> The final result is

$$\langle \delta_g(\vec{k}) \delta_g(\vec{K} - \vec{k}) \rangle_{\sim \vec{K} \text{ fixed}} = (2\pi)^3 \delta_D(\vec{K}) b_1^2 P_{\text{lin}}(k) + b_1 \sum_{\alpha} c_{\alpha} f_{\alpha}(\vec{k}, \vec{K} - \vec{k}) \delta_1(\vec{K}), \quad (5.9)$$

where

$$f_{\alpha}(\vec{k}_1, \vec{k}_2) \equiv 2[F_{\alpha}(\vec{k}_1 + \vec{k}_2, -\vec{k}_1) P_{\text{lin}}(|\vec{k}_1|, z) + 1 \leftrightarrow 2]. \quad (5.10)$$

In Equation (5.9), we find the same structure as in the CMB lensing case in Eq. (5.6): the standard power spectrum term, plus a term from off-diagonal correlations induced by the fixed background mode.

Equation (5.9) suggests we can multiply two different modes of the measured tracer field and then simply “divide” by the coupling strength  $b_1 \sum_{\alpha} c_{\alpha} f_{\alpha}(\vec{k}, \vec{K} - \vec{k})$  to obtain an estimate of the linear field  $\delta_1$  at large scales. Unfortunately, in general we do not know the bias coefficients  $b_1$  or  $c_{\alpha}$  a priori, so the best we can do is to use the galaxy mode couplings to estimate the product  $b_1 c_{\alpha} \delta_1$  for a chosen  $\alpha$ . To reduce variance on the estimate, we will sum over all the mode couplings that involve the same large-scale mode. This can be achieved by writing the following general quadratic estimator

$$\hat{\Delta}_{\alpha}(\vec{K}) \equiv \widehat{b_1 c_{\alpha} \delta_1}(\vec{K}) = \int_{\vec{q}} g_{\alpha}(\vec{q}, \vec{K} - \vec{q}) \delta_g(\vec{q}) \delta_g(\vec{K} - \vec{q}), \quad (5.11)$$

with weights  $g_{\alpha}$ , similar to what is done for CMB lensing [128] or “clustering fossils” from primordial gravitational waves [173, 137, 174]. For an alternative derivation of this estimator, based on optimizing the cross-correlation of a quadratic combination of measured modes with the true linear mode to be reconstructed, see Appendix C.1.

<sup>4</sup>One must integrate in a neighborhood around the argument of a Dirac delta function for this collapse to take place, and this is why we considered a neighborhood around  $\vec{K}$  in the first place. In the discrete case, where we have sums instead of integrals over wavenumbers, we could define our ensemble average to keep a single mode  $\delta_1(\vec{K})$  fixed, since we would then have Kronecker deltas instead of Dirac delta functions.

The covariance between two such estimators  $\alpha$  and  $\beta$  of the biased matter density field on large scales can be split into a Gaussian part, coming from all disconnected contributions, and a non-Gaussian part that includes all connected contributions:

$$\begin{aligned} & \langle \hat{\Delta}_\alpha(\vec{K}) \hat{\Delta}_\beta^*(\vec{K}') \rangle - \langle \hat{\Delta}_\alpha(\vec{K}) \rangle \langle \hat{\Delta}_\beta^*(\vec{K}') \rangle \\ &= (2\pi)^3 \delta_D(\vec{K} - \vec{K}') \left[ \text{Cov}_G(\hat{\Delta}_\alpha(\vec{K}), \hat{\Delta}_\beta^*(\vec{K}')) + \text{Cov}_{NG}(\hat{\Delta}_\alpha(\vec{K}), \hat{\Delta}_\beta^*(\vec{K}')) \right]. \end{aligned} \quad (5.12)$$

We constrain the weights to provide an estimator that is optimal in the sense of minimizing the Gaussian contributions to its variance,

$$\text{Var}_G[\hat{\Delta}_\alpha](\vec{K}) \equiv \text{Cov}_G(\hat{\Delta}_\alpha(\vec{K}), \hat{\Delta}_\alpha^*(\vec{K}')) , \quad (5.13)$$

while requiring that it be unbiased if there were only a single mode-coupling, i.e.

$$\int_{\vec{q}} g_\alpha(\vec{q}, \vec{K} - \vec{q}) f_\alpha(\vec{q}, \vec{K} - \vec{q}) = 1 . \quad (5.14)$$

These criteria lead to the familiar quadratic estimator weights:

$$g_\alpha(\vec{k}_1, \vec{k}_2) = N_{\alpha\alpha}(\vec{k}_1 + \vec{k}_2) \frac{f_\alpha(\vec{k}_1, \vec{k}_2)}{2P_{\text{tot}}(k_1)P_{\text{tot}}(k_2)} , \quad (5.15)$$

where  $P_{\text{tot}}$  is the sum of the clustering and shot noise contributions to the tracer power spectrum. The normalization is given by

$$N_{\alpha\beta}(\vec{K}) = \left( \int_{\vec{q}} \frac{f_\alpha(\vec{q}, \vec{K} - \vec{q}) f_\beta(\vec{q}, \vec{K} - \vec{q})}{2P_{\text{tot}}(q)P_{\text{tot}}(|\vec{K} - \vec{q}|)} \right)^{-1} \quad (5.16)$$

which guarantees that  $N_{\alpha\alpha}$  is equal to the Gaussian part of the variance of  $\hat{\Delta}_\alpha$ . We will refer to  $N_{\alpha\alpha}$  as the reconstruction noise, which incorporates cosmic variance in the reconstruction and the disconnected contribution from shot noise of the tracer field. (It should be noted that  $N_{\alpha\beta}$  is not equal to the noise when  $\alpha \neq \beta$ .) With the weights in Eq. (5.15), the estimator in Eq. (5.11) becomes

$$\hat{\Delta}_\alpha(\vec{K}) = N_{\alpha\alpha}(\vec{K}) \int_{\vec{q}} \frac{f_\alpha(\vec{q}, \vec{K} - \vec{q})}{2P_{\text{tot}}(q)P_{\text{tot}}(|\vec{K} - \vec{q}|)} \delta_g(\vec{q}) \delta_g(\vec{K} - \vec{q}) . \quad (5.17)$$

The non-Gaussian part of the variance includes a trispectrum contribution from clustering of the tracers, and further contributions from tracer shot noise. We neglect the former, because it is subdominant to the latter; our comparisons with simulations in Sec. 5.3 show that this is a valid approximation. Importantly, the shot noise

contributions can dominate over the Gaussian reconstruction noise in many cases, because these contributions couple to large-scale modes with large variance, while the Gaussian contribution only involves small-scale modes, which have smaller variance due to the shape of the matter power spectrum. We derive the full expressions for these contributions and discuss their hierarchy further in Appendix C.2.

The expectation value of the estimator in Eq. (5.17), for a given realization of the linear field at wavevector  $\vec{K}$ , is

$$\langle \hat{\Delta}_\alpha(\vec{K}) \rangle_{\delta_1(\vec{K}) \text{ fixed}} = b_1 \left[ c_\alpha + \sum_{\beta \neq \alpha} c_\beta \frac{N_{\alpha\alpha}(K)}{N_{\alpha\beta}(K)} \right] \delta_1(\vec{K}) . \quad (5.18)$$

We clearly see that there is a contamination of the estimator with respect to the case of only a single mode-coupling, given by the product of the (Gaussian) noise for the estimator  $\alpha$  and a sum of bias terms divided by the cross normalization between estimators  $\alpha$  and  $\beta$ .<sup>5</sup> If the goal is to just reconstruct the linear mode of interest, then it is important to account for this contribution. One can attempt to construct a so-called ‘‘bias-hardened’’ estimator by forming a linear combination of the original estimators that is free of this contamination at leading order (e.g. [197, 201, 91]). However, for the specific mode-couplings relevant in this situation, the high degree of correlation between the original estimators implies that the noise on the new estimator will be so high that it is no longer useful; see Appendix C.3 for details.

We claim that, for extracting non-Gaussianity, this contamination can actually be useful. As we will see later, some of these contaminating terms induce scale-dependence that reproduces the  $1/K^2$  scaling created by primordial non-Gaussianity. Depending of the signs of these terms, they can either raise or lower the signal to noise on  $f_{\text{NL}}$  from the reconstructed field. We will discuss this further in Sec. 5.2.3.

## 5.2.2 Non-Gaussianity and bias expansion

As we discussed in the introduction, primordial non-Gaussianity of the local type introduces a quadratic contribution to the metric perturbation. The metric perturbation (gravitational potential)  $\varphi$  is related to the linear matter overdensity through the usual Poisson equation (dropping the subscript  $\text{G}$ )

$$\varphi(\vec{k}, z) = \frac{\delta_1(\vec{k}, z)}{M(k, z)}, \quad (5.19)$$

<sup>5</sup>It can be seen from Eq. (5.16) that as the overlap integral of the two mode-couplings goes to zero,  $N_{\alpha\beta}$  becomes very large and the contamination vanishes.

where the Poisson factor  $M(k, z)$  is given by

$$M(k, z) = \frac{2c^2}{3H_0^2\Omega_m} D(z)k^2T(k). \quad (5.20)$$

Here, the growth factor  $D(z)$  is normalized to agree with the scale factor  $1/(1+z)$  during matter domination.

Galaxies and 21 cm fluctuations of the density field are biased tracers of the underlying, dynamically dominant matter distribution. In the presence of local primordial non-Gaussianity, the coupling of the long and short modes leads to an additional modulation of the abundance of collapsed objects by the long wavelength potential fluctuations  $\varphi$ . To describe biased tracers we thus follow [97, 23] in performing a double expansion of the Eulerian galaxy (or tracer) density field in the non-linear density and linear potential<sup>6</sup>

$$\begin{aligned} \delta_g^E(\vec{x}) = & b_{10}^E \delta(\vec{x}) + b_{01}^E \varphi(\vec{x}_L[\vec{x}]) + b_{20}^E \delta^2(\vec{x}) + b_{11}^E \delta(\vec{x}) \varphi(\vec{x}_L[\vec{x}]) + b_{02}^E \varphi^2(\vec{x}) \\ & + b_{s^2}^E s_{ij}(\vec{x}) s^{ij}(\vec{x}) + \varepsilon(\vec{x}) + \varepsilon_\delta(\vec{x}) \delta(\vec{x}) + \varepsilon_\varphi(\vec{x}) \varphi(\vec{x}_L[\vec{x}]) + \dots \end{aligned} \quad (5.21)$$

Here, the  $b_{ij}^E$  are the Eulerian bias parameters<sup>7</sup>,  $s_{ij}$  is the tidal tensor

$$s_{ij}(\vec{x}) = \left[ \frac{\nabla_i \nabla_j}{\nabla^2} - \frac{1}{3} \delta_{ij}^{(K)} \right] \delta(\vec{x}), \quad (5.22)$$

and  $\varepsilon$  is the stochasticity, which correlates with itself but not with the linear density field. In the simplest case where galaxies are a Poisson sample of the underlying matter field, the stochasticity leads to the fiducial  $1/\bar{n}$  power spectrum. The higher order stochasticity contributions  $\varepsilon_\delta \delta$  and  $\varepsilon_\varphi \varphi$  lead to stochasticity contributions in the bispectrum [79], as we review in App. C.2. In simple local-Lagrangian bias models, the tidal tensor bias  $b_{s^2}^E$  can be related to the linear density bias as  $b_{s^2}^E = -2/7 (b_{10}^E - 1)$  [24]. Employing realistic simplifying assumptions, we will see that all of the bias parameters  $b_{ij}^E$  can be expressed in terms of  $b_{10}^E$  and  $b_{20}^E$ . We are truncating the above expansion at second order, since we will only consider tree level power spectra and bispectra as well as the Gaussian disconnected trispectrum in our derivations. We can thus also neglect higher derivative contributions, such as  $k^2 \delta_1(\vec{k})$ , as the contribution of their bias parameters are equivalent to cubic contributions to the matter and galaxy

<sup>6</sup>In the peak-background split formalism, the abundance of collapsed objects is given by  $e^{-\nu^2/2}$  where  $\nu = \delta_c/\sigma$  with  $\delta_c$  the collapse threshold and  $\sigma$  the variance. The long wavelength density modulates the collapse threshold as  $\delta_c \rightarrow \delta_c - \delta$ , whereas the metric perturbation  $\varphi$  modulates the variance  $\sigma \rightarrow \sigma(1 + 2f_{\text{NL}}\varphi)$ .

<sup>7</sup>In the introductory discussion in Sec. 5.2.1, we employed the notation  $b_1 \equiv b_{10}^E$  for the sake of simplicity.

density fields (e.g. [1]). Note that all of the  $\delta$  terms in Eq. (5.21) refer to the underlying non-linear matter density field including its quadratic couplings. The potential  $\varphi$ , in turn, is linear, as the dependence of the halo abundance on long wavelength potential fluctuations is set up in the early Universe.

There is, however, a non-linearity in the potential terms that arises from the fact that the abundance of galaxies in the peak-background split is set up in Lagrangian space with coordinates  $\vec{x}_L$ . These Lagrangian positions are related to the Eulerian coordinates by  $\vec{x}_L[\vec{x}] = \vec{x} - \vec{\Psi}(\vec{x})$  at leading order. The potential is thus advected by long wavelength displacements as [274]

$$\varphi(\vec{x}_L[\vec{x}]) = \varphi(\vec{x}) - \vec{\Psi}(\vec{x}) \cdot \vec{\nabla} \varphi(\vec{x}) + \dots \quad (5.23)$$

The Fourier transform of the linear displacement field  $\vec{\Psi}(\vec{x})$  is related to the linear matter overdensity by  $\vec{\Psi}(\vec{k}) = i(\vec{k}/k^2)\delta(\vec{k})$ .

At second order, the matter density field picks up a new quadratic contribution from primordial non-Gaussianity according to Eq. (5.2):

$$\delta(\vec{k}) = \delta_1(\vec{k}) + \int_{\vec{q}} \left[ \sum_{\alpha=G,S,T} F_\alpha(\vec{q}, \vec{k} - \vec{q}) \right] \delta_1(\vec{q}) \delta_1(\vec{k} - \vec{q}) + f_{\text{NL}} M(k) \int_{\vec{q}} \varphi(\vec{q}) \varphi(\vec{k} - \vec{q}) + \dots, \quad (5.24)$$

where the growth, shift and tidal components of the gravitational coupling kernel are given by Eq. (5.5). For biased tracers, this expression gets multiplied by  $b_{10}^E$ . We can rewrite the last term in terms of the density field using the Poisson equation, resulting in a new quadratic coupling

$$F_{\varphi\varphi}(\vec{k}_1, \vec{k}_2) = \frac{M(|\vec{k}_1 + \vec{k}_2|)}{M(k_1)M(k_2)}, \quad (5.25)$$

such that

$$\delta(\vec{k}) = \delta_1(\vec{k}) + \int_{\vec{q}} \left[ \sum_{\alpha=G,S,T,\varphi\varphi} c_\alpha f_{\text{NL}}^{p_\alpha} F_\alpha(\vec{q}, \vec{k} - \vec{q}) \right] \delta_1(\vec{q}) \delta_1(\vec{k} - \vec{q}) + \dots, \quad (5.26)$$

where now  $c_\alpha = \{1, 1, 1, 1\}$  and  $p_\alpha = \{0, 0, 0, 1\}$ .

Combining this result with the Fourier transform of the other second order bias terms in Eq. (5.21) yields

$$\begin{aligned}
\delta_{\mathbf{g}}^{\text{E}}(\vec{k}) = & \left[ b_{10}^{\text{E}} + \frac{b_{01}^{\text{E}}}{M(k)} \right] \delta_1(\vec{k}) + b_{01}^{\text{E}} \int_{\vec{q}} \frac{1}{2} \left[ \frac{\vec{q} \cdot (\vec{k} - \vec{q})}{q^2 M(|\vec{k} - \vec{q}|)} + \frac{\vec{q} \cdot (\vec{k} - \vec{q})}{|\vec{k} - \vec{q}|^2 M(q)} \right] \delta_1(\vec{q}) \delta_1(\vec{k} - \vec{q}) \\
& + b_{10}^{\text{E}} \int_{\vec{q}} \left[ \sum_{\alpha=\text{G,S,T},\varphi\varphi} F_{\alpha}(\vec{q}, \vec{k} - \vec{q}) \right] \delta_1(\vec{q}) \delta_1(\vec{k} - \vec{q}) \\
& + f_{\text{NL}} b_{10}^{\text{E}} \int_{\vec{q}} \frac{M(k)}{M(q)M(|\vec{k} - \vec{q}|)} \delta_1(\vec{q}) \delta_1(\vec{k} - \vec{q}) + b_{20}^{\text{E}} \int_{\vec{q}} \delta_1(\vec{q}) \delta_1(\vec{k} - \vec{q}) \\
& + b_{11}^{\text{E}} \int_{\vec{q}} \frac{1}{2} \left( \frac{1}{M(q)} + \frac{1}{M(|\vec{k} - \vec{q}|)} \right) \delta_1(\vec{q}) \delta_1(\vec{k} - \vec{q}) \\
& + b_{02}^{\text{E}} \int_{\vec{q}} \frac{1}{M(q)M(|\vec{k} - \vec{q}|)} \delta_1(\vec{q}) \delta_1(\vec{k} - \vec{q}) \\
& + b_{s^2}^{\text{E}} \int_{\vec{q}} \left[ \frac{[\vec{q} \cdot (\vec{k} - \vec{q})]^2}{q^2 |\vec{k} - \vec{q}|^2} - \frac{1}{3} \right] \delta_1(\vec{q}) \delta_1(\vec{k} - \vec{q}). \tag{5.27}
\end{aligned}$$

The additional terms arising from the non-Gaussian bias can be encoded by the new quadratic coupling kernels

$$\begin{aligned}
F_{01} = \frac{1}{2} \vec{k}_1 \cdot \vec{k}_2 \left( \frac{1}{k_2^2 M(k_1)} + \frac{1}{k_1^2 M(k_2)} \right), \quad F_{11} = \frac{1}{2} \left( \frac{1}{M(k_1)} + \frac{1}{M(k_2)} \right), \\
F_{02} = \frac{1}{M(k_1)M(k_2)}. \tag{5.28}
\end{aligned}$$

The Eulerian bias parameters can be related to their Lagrangian counterparts through a spherical collapse calculation [97, 23]:

$$b_{10}^{\text{E}} = b_{10}^{\text{L}} + 1, \tag{5.29}$$

$$b_{20}^{\text{E}} = 2(a_1 + a_2)b_{10}^{\text{L}} + a_1^2 b_{20}^{\text{L}}, \tag{5.30}$$

$$b_{01}^{\text{E}} = b_{01}^{\text{L}}, \tag{5.31}$$

$$b_{11}^{\text{E}} = a_1 b_{11}^{\text{L}} + b_{01}^{\text{L}}, \tag{5.32}$$

$$b_{02}^{\text{E}} = b_{02}^{\text{L}}, \tag{5.33}$$

where  $a_1 = 1$  and  $a_2 = -17/21$  are spherical collapse expansion factors. The non-Gaussian Lagrangian bias parameters can be obtained using the peak background split. They are given as the derivatives of the mass function with respect to the long wavelength potential fluctuations. Assuming a universal mass function, the derivatives with respect to the potential can be related to the derivatives with respect to the long

Mode Coupling ( $\alpha$ )	$p_\alpha$	$c_\alpha$	$F_\alpha(\vec{k}_1, \vec{k}_2)$
G	0	$b_1 + \frac{21}{17}b_2$	$\frac{17}{21}$
S	0	$b_1$	$\frac{1}{2} \left[ \frac{1}{k_1^2} + \frac{1}{k_2^2} \right] (\vec{k}_1 \cdot \vec{k}_2)$
T	0	$b_1 + \frac{7}{2}b_{s^2}$	$\frac{2}{7} \left[ \frac{(\vec{k}_1, \vec{k}_2)^2}{k_1^2 k_2^2} - \frac{1}{3} \right]$
$\varphi\varphi$	1	$b_1$	$\frac{M( \vec{k}_1 + \vec{k}_2 , z)}{M(k_1)M(k_2)}$
01	1	$2\delta_c(b_1 - 1)$	$\frac{1}{2} \vec{k}_1 \cdot \vec{k}_2 \left( \frac{1}{k_2^2 M(k_1)} + \frac{1}{k_1^2 M(k_2)} \right)$
11	1	$2 \left( \delta_c \left[ \frac{b_2 - 2(a_1 + a_2)(b_1 - 1)}{a_1} \right] - a_1 [b_1 - 1] \right) + 2\delta_c (b_1 - 1)$	$\frac{1}{2} \left( \frac{1}{M(k_1)} + \frac{1}{M(k_2)} \right)$
02	2	$4\delta_c \left( \delta_c \left[ \frac{b_2 - 2(a_1 + a_2)(b_1 - 1)}{a_1^2} \right] - 2[b_1 - 1] \right)$	$\frac{1}{M(k_1)M(k_2)}$

Table 5.2 Mode couplings,  $f_{\text{NL}}$  exponents, bias parameters and coupling kernels of the quadratic interactions for Eq. (5.37).

wavelength density, and consequently the bias parameters of the potential terms can be related to the bias parameters of the density terms:

$$b_{01}^L = 2f_{\text{NL}}\delta_c \left( b_{10}^E - 1 \right) , \quad (5.34)$$

$$b_{11}^L = 2f_{\text{NL}} \left( \delta_c \left[ \frac{b_{20}^E - 2(a_1 + a_2)(b_{10}^E - 1)}{a_1^2} \right] - [b_{10}^E - 1] \right) , \quad (5.35)$$

$$b_{02}^L = 4f_{\text{NL}}^2\delta_c \left( \delta_c \left[ \frac{b_{20}^E - 2(a_1 + a_2)(b_{10}^E - 1)}{a_1^2} \right] - 2[b_{10}^E - 1] \right) , \quad (5.36)$$

where  $\delta_c$  is the spherical collapse threshold. Note that small deviations from this simple scaling of non-Gaussian bias  $b_{01}^L$  with Gaussian bias  $b_{10}^E$  have been found in simulations [40] and seem to depend on the way halos are identified.

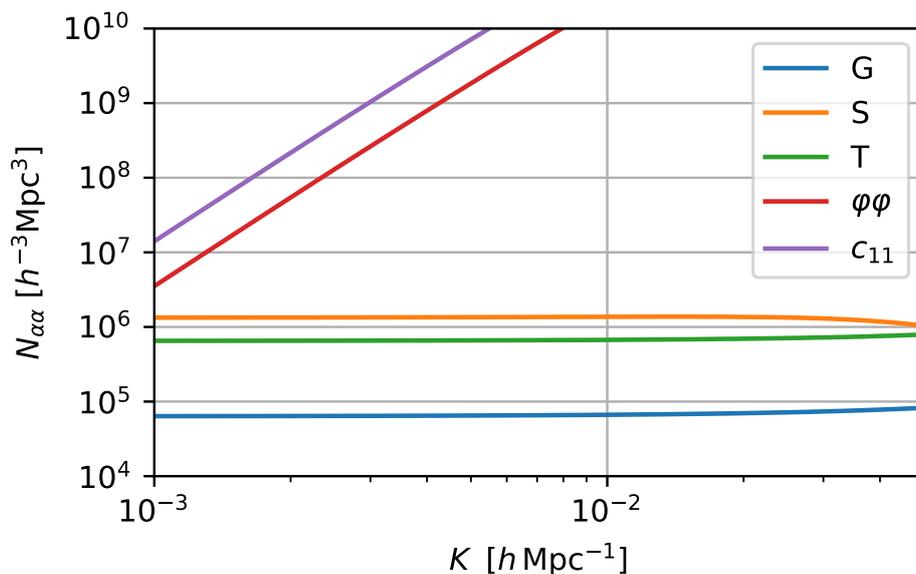
In summary, we can write for the galaxy density field up to second order in the presence of local type primordial non-Gaussianity:

$$\delta_g(\vec{k}) = \left[ b_{10}^E + f_{\text{NL}} \frac{c_{01}}{M(k)} \right] \delta_1(\vec{k}) + \int_{\vec{q}} \left[ \sum_{\alpha} c_{\alpha} f_{\text{NL}}^{p_{\alpha}} F_{\alpha}(\vec{q}, \vec{k} - \vec{q}) \right] \delta_1(\vec{q}) \delta_1(\vec{k} - \vec{q}) , \quad (5.37)$$

where  $\alpha$  now runs over  $\{\text{G}, \text{S}, \text{T}, \varphi\varphi, 01, 11, 02\}$  with the couplings given in Table 5.2. In this table, Eq. (5.37), and throughout the rest of the chapter, we have simplified the notation to  $b_1 \equiv b_{10}^E$ ,  $b_2 \equiv b_{20}^E$ , and  $b_{s^2} \equiv b_{s^2}^E$ . Note that we have not included mode-couplings due to lensing, which are expected to be a subdominant contribution that is somewhat degenerate with the S term [91], nor have we incorporated redshift space distortions or anisotropic selection effects (see Sec. 5.4.3 for discussion).

### 5.2.3 Reconstruction noise and contamination

With this formalism in place, we can now examine the noise of the reconstructed modes, and the contamination arising from the presence of multiple mode-couplings in the tracer field used for reconstruction.<sup>8</sup> We will show these quantities for a DESI-like survey (with specifications given in Sec. 5.4.3), but we have checked that the conclusions we draw from this case also apply to the other surveys we consider.



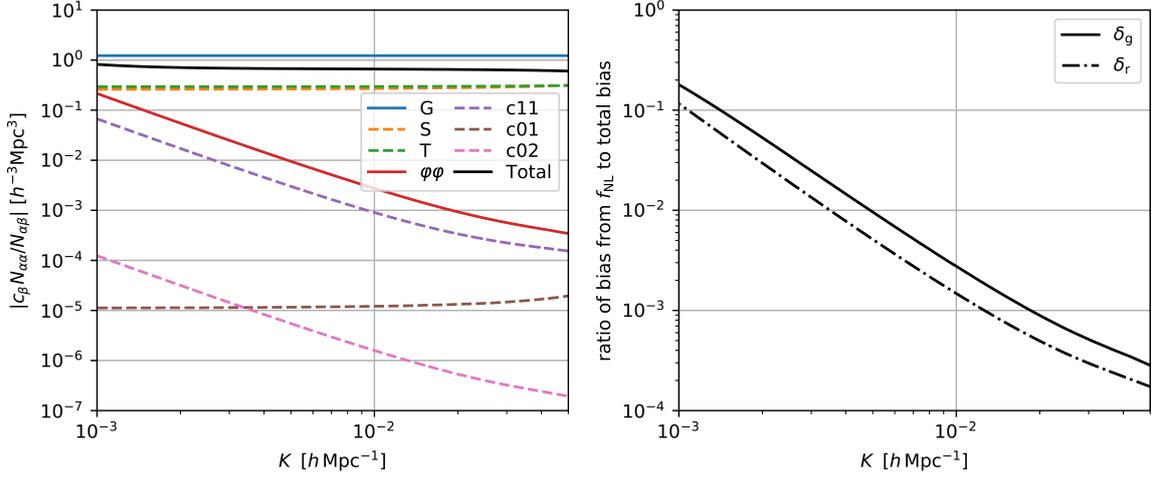
*Fig. 5.1* Reconstruction noise power spectra for estimators that use each of the quadratic mode-couplings discussed in Sec. 5.2.2. We omit curves for the  $c_{01}$  and  $c_{02}$  estimators, which are greater than the upper limit of the plot. The G (“growth”) estimator has the lowest noise by far. These curves are computed for a DESI-like survey, but the hierarchy between them is unchanged for the other surveys we consider. The signal to noise on reconstructed modes (not shown) is likewise much higher for the G estimator than for S or T, justifying our use of the G estimator for our main results.

Figure 5.1 shows the reconstruction noise power spectrum corresponding to estimators that use each of the quadratic mode-couplings discussed in Sec. 5.2.2. We see that the “growth” estimator has the lowest noise by far. We compare the predicted noise for the G, S, and T estimators with results from  $N$ -body simulations in Sec. 5.3 (among other tests), finding good agreement. Thus, we use the growth estimator in our forecasts for reconstruction<sup>9</sup>, henceforth referring to reconstructed modes as  $\delta_r(\vec{K})$

<sup>8</sup>For producing matter power spectra for forecasts, we relied on the `nbodykit` code (<https://github.com/bccp/nbodykit>).

<sup>9</sup>Out of the G, S, and T estimators, the G estimator yields both the lowest noise and the highest signal to noise on reconstructed modes. However, some of the other estimators (e.g.  $\alpha = \varphi\varphi$ ) also have

instead of  $\hat{\Delta}_G(\vec{K})$ . However, as we discussed in Sec. 5.2.1, the output of the G estimator (or any other single estimator) will be contaminated by the other mode-couplings, with the specific contamination given by Eq. (5.18), and we must incorporate this contamination into our forecasts.



*Fig. 5.2 Left:* Contamination in the expectation value of G estimator, corresponding to separate multiplicative biases on the amplitude of a reconstructed mode, computed for a DESI-like survey. Blue solid line is the estimator growth bias shown for comparison. Dashed lines indicate negative values. Several of these curves inherit the  $k^{-2}$  scaling of the scale-dependent bias in  $\delta_g$  arising from nonzero  $f_{\text{NL}}$ , implying that reconstructed modes can be used to constrain  $f_{\text{NL}}$  in the same way. *Right:* Ratio of scale-dependent bias from  $f_{\text{NL}}$  (for a fiducial value of  $f_{\text{NL}} = 1$ ) to total bias for  $\delta_g$  (solid) and  $\delta_r$  (dot-dashed). Local primordial non-Gaussianity has roughly the same relative contribution to the bias of  $\delta_g$  or reconstructed modes.

We show this contamination in the left panel of Figure 5.2, in the form of each term  $c_\beta N_{\text{GG}}/N_{\text{G}\beta}$  in the square brackets of Eq. (5.18). These curves each represent separate multiplicative biases on the amplitude of a reconstructed mode. Those arising from late-time gravitational evolution (S, T) or from advection of the primordial potential ( $c_{01}$ ) are white in  $K$ . In contrast, those arising from couplings between  $\delta$  and  $\varphi$  ( $c_{11}$ ) or  $\varphi$  and itself ( $\varphi\varphi$ ,  $c_{02}$ ) scale like  $M(K)^{-1} \propto K^{-2}$ . We derive these scalings analytically in the large-scale limit in Appendix C.4. Importantly, all terms that scale like  $K^{-2}$  involve  $f_{\text{NL}}$ , such that, as for  $\delta_g$ , low- $K$  scale-dependent bias in the reconstructed modes can be used as a probe of local primordial non-Gaussianity. The right panel of Fig. 5.2 shows that the relative size of this scale-dependent bias is comparable for  $\delta_g$  and  $\delta_r$ , reaching  $\mathcal{O}(10\%)$  at  $K \sim 0.001 h \text{ Mpc}^{-1}$ , assuming  $f_{\text{NL}} = 1$ .

signal to noise approaching that of the G estimator, since the contaminating terms in Eq. (5.18) act as “signal” in a signal-to-noise computation. This indicates that a more optimal choice of estimator weights may be possible, although we leave this to future work.

Figure 5.2 also shows that the contamination from other mode-couplings is subdominant to the intrinsic bias on the reconstructed field (i.e. the  $c_G$  term in Eq. 5.18). Thus, using  $c_G = b_1 + (21/17)b_2$  from Table 5.2, we can derive the rough dependence of  $P_{\text{rr}}$  and  $P_{\text{gr}}$  on  $b_1$  and  $b_2$ :

$$P_{\text{rr}} \propto b_1^2(b_1 + b_2)^2, \quad P_{\text{gr}} \propto b_1^2(b_1 + b_2). \quad (5.38)$$

If galaxy shot noise is negligible compared to  $P_{\text{gg}}$ , then the reconstruction noise  $N_{\text{GG}}$  satisfies  $N_{\text{GG}} \propto b_1^4$ , implying that  $P_{\text{rr}}/N_{\text{GG}} \propto (1 + b_2/b_1)^2$  in this regime. This scaling will be useful to help understand the behavior of our forecasts when we change the fiducial value of  $b_2$ .

## 5.3 Simulations

To validate the quadratic estimator framework presented in Sec. 5.2, we use a suite of 15 realisations of a cosmological  $N$ -body simulation [1]. The initial conditions are generated with the second-order Lagrangian Perturbation Theory (**2-LPT**) code [241] at the initial redshift  $z_i = 99$  and are subsequently evolved using **Gadget-2** [266] for  $f_{\text{NL}} = 0$ . The simulations are performed with  $N_p = 1024^3$  dark matter particles in a cubic box of length  $L = 1500h^{-1}$  Mpc with periodic boundary conditions. We assume a flat  $\Lambda$ CDM cosmology with the cosmological parameters  $\Omega_m = 0.272$ ,  $\Omega_\Lambda = 0.728$ ,  $h = 0.704$ ,  $n_s = 0.967$ ,  $\sigma_8 = 0.81$ .

We restrict our simulation comparisons to  $z = 0$  because this is where nonlinearities are the strongest and our quadratic bias approximation breaks down at the lowest wavenumber; analyzing a field at  $z = 0$  is therefore, arguably, the most rigorous test of our methodology. Since the purpose of the simulation comparisons is to provide a first, simple validation of our reconstruction procedure, rather than to fully explore its simulated performance as a function of redshift in detail, we defer any detailed explorations beyond our  $z = 0$  validation to future work.

Dark matter halos in the final  $z = 0$  density field are identified using a Friends-of-Friends (FoF) algorithm with linking length  $l = 0.2$  times the mean interparticle distance. The halos are binned in mass, with each bin spanning a factor of three in mass. We have checked the viability of our reconstruction method for a range of masses, finding qualitatively similar results in all cases; however, for simplicity, we present only the results for the lowest mass bin, the properties of which are given in Table 5.3. Particles and halos are assigned to a regular grid using the Cloud-in-Cell

Mass Bin	Mean Halo Mass [ $10^{13}h^{-1}M_{\odot}$ ]	$\bar{n}$ [ $10^{-6}h^3 \text{ Mpc}^{-3}$ ]	$b_1$	$c_g$	$c_t$	$c_s$
I	0.77	627	1.07	0.62	1.14	1.07

*Table 5.3* Properties of the halo mass bin employed in this study: the mean mass of the sample, the number density of halos  $\bar{n}$ , the linear bias  $b_1$ , and the three relevant  $c_{\alpha}$  parameters defined in Table 5.2. The measured bias parameters are taken from [1], which is based on the same simulations and mass bin we use here.

(CIC) scheme. We Fourier transform the matter and halo density fields using the publicly available **FFTW** library<sup>10</sup>.

### 5.3.1 Generation of quadratic estimators

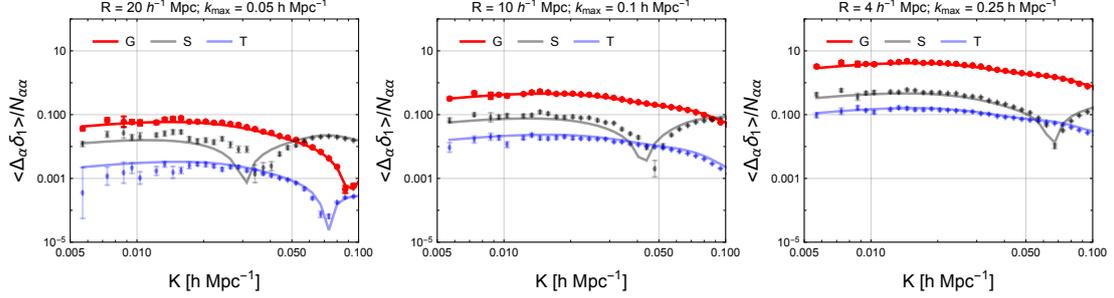
We generate quadratic estimators from the halo density field  $\delta_g$  in  $N$ -body simulations using the convolution theorem. This means that we use a sequence of multiplications with powers of wavenumbers in Fourier space, Fourier transforms, and subsequent multiplication of the weighted fields in configuration space. We generate three quadratic estimators corresponding to the growth term  $\delta^2$ , shift term  $\Psi \cdot \nabla \delta$ , and the tidal term  $s^2$ , with associated Fourier-space kernels given in Eq. (5.5). The first step in our procedure is to remove very small scale modes by applying a cut-off  $k_{\max}$  in Fourier space through multiplication of the Fourier space density field with a filtering function. While the exact form of the cutoff is not important, we adopt a Gaussian filter  $W(R\vec{k}) = \exp(-k^2 R^2/2)$  for numerical stability. We define the smoothed density field by  $\delta_g^R(\vec{k})$ . We choose three external smoothing scales:  $R = 20h^{-1}$  Mpc,  $R = 10h^{-1}$  Mpc, and  $R = 4h^{-1}$  Mpc, corresponding to maximum wavenumbers  $k_{\max} \approx 0.05h \text{ Mpc}^{-1}$ ,  $k_{\max} \approx 0.1h \text{ Mpc}^{-1}$ , and  $k_{\max} \approx 0.25h \text{ Mpc}^{-1}$  respectively. The smoothing scale removes all wavenumbers  $k > k_{\max}$ , such that we reconstruct long wavelength modes using modes  $k < k_{\max}$  for three different cases.

The mode coupling functions  $g_{\alpha}(\vec{q}, \vec{k} - \vec{q})$  defined by Eq. (5.15) contain a Wiener filter, which we implement by first generating the linear power spectrum on the simulation grid, and then defining two fields:

$$\delta_A(\vec{k}) = \frac{\delta_g^R(\vec{k})}{b_1^2 P_{\text{lin}}(\vec{k}) + \bar{n}^{-1}} \quad \text{and} \quad \delta_B(\vec{k}) = \frac{\delta_g^R(\vec{k}) P_{\text{lin}}(\vec{k})}{b_1^2 P_{\text{lin}}(\vec{k}) + \bar{n}^{-1}}, \quad (5.39)$$

where  $b_1$  and  $\bar{n}$  are the linear bias and halo number density corresponding to the halo mass bin defined in Table 5.3. Using  $\delta_A$  and  $\delta_B$  we generate growth, shift and tidal estimators using multiplications of powers of wavenumbers in Fourier space, Fourier

<sup>10</sup><http://www.fftw.org>



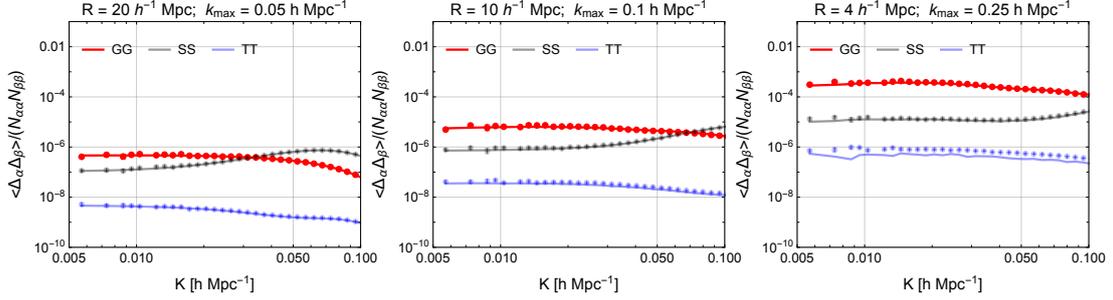
*Fig. 5.3* Cross correlations of estimators  $\hat{\Delta}_\alpha$  corresponding to the growth, shift, and tidal mode-couplings with the linear density field  $\delta_1$ . We compare theory predictions (lines) with simulations (points) for three different smoothing scales,  $R = 20h^{-1}$  Mpc,  $R = 10h^{-1}$  Mpc and  $R = 4h^{-1}$  Mpc, corresponding to maximum wavenumbers  $k_{\max} = 0.05h \text{ Mpc}^{-1}$ ,  $0.1h \text{ Mpc}^{-1}$ , and  $0.25h \text{ Mpc}^{-1}$  respectively. In this figure, we plot  $\langle \hat{\Delta}_\alpha \delta_1 \rangle / N_{\alpha\alpha}$  (in contrast to what is defined in Eq. (5.17), in simulations we define the estimators  $\hat{\Delta}_\alpha$  without a prefactor  $N_{\alpha\alpha}$ ). We find very good agreement for the growth estimator for all smoothing scales, and also reasonably good agreement for the other estimators.

transforms, and multiplication of fields in configurations space. For example, we generate the growth estimator as follows. First, we inverse Fourier transform both fields defined in Eq. (5.39) to obtain  $\delta_A(\vec{x})$  and  $\delta_B(\vec{x})$ . Next, in configuration space, we multiply the product of both fields by  $17/21$  (Table 5.2) and finally Fourier transform back to obtain the growth estimator in Fourier space. We generate shift and tidal estimators with a similar procedure.

Note that the main computational cost in generating the quadratic estimators comes from performing the Fourier transforms. The auto- and cross-spectrum analysis of quadratic estimators only requires the computational cost of a power spectrum analysis, which is quite efficient. In all our figures in this section, we estimate the errorbars of our measurements using the standard deviation of 15 simulation realisations.

### 5.3.2 Cross-correlation of quadratic estimators with the initial linear field

In this section, we describe our results for the cross-correlations of three quadratic estimators  $\hat{\Delta}_\alpha(\vec{k})$  with the initial linear field  $\delta_1(\vec{k})$ , and compare the theory predictions with simulations. The prediction is given by (since  $f_{\text{NL}} = 0$  we only consider the



*Fig. 5.4* Auto-correlations of the quadratic estimators  $\hat{\Delta}_\alpha$ , for the same smoothing scales shown in Fig. 5.3. The predictions for the growth estimator agree with simulations for all smoothing scales. However, for other estimators predictions agree with simulations for large smoothing scales but for the low smoothing scales, the predictions slightly disagree with simulation results as higher-order terms become more important.

growth, shift and tidal terms)

$$\begin{aligned}
 \langle \hat{\Delta}_\alpha(\vec{k}) \delta_1(\vec{k}') \rangle' &= b_1 N_{\alpha\alpha}(\vec{k}) \sum_{\beta \in \{G, S, T\}} c_\beta P_{\text{lin}}(\vec{k}) \\
 &\times \int_{\vec{q}} \frac{f_\alpha(\vec{q}, \vec{k} - \vec{q}) f_\beta(\vec{q}, \vec{k} - \vec{q})}{2P_{\text{tot}}(\vec{q}) P_{\text{tot}}(\vec{k} - \vec{q})} W(R\vec{q}) W(R(\vec{k} - \vec{q})) + P_{\alpha, \text{shot}}(\vec{k}) \quad (5.40) \\
 &= b_1 P_{\text{lin}}(\vec{k}) \sum_{\beta \in \{G, S, T\}} c_\beta \frac{N_{\alpha\alpha}(\vec{k})}{N_{\alpha\beta}(\vec{k})} + P_{\alpha, \text{shot}}(\vec{k}) ,
 \end{aligned}$$

where the prime on the left-hand side denotes that the factor of  $(2\pi)^3 \delta_{\text{D}}(\vec{k} + \vec{k}')$  has been omitted, and  $c_\beta$  are bias parameters corresponding to the growth, shift and tidal terms and can be measured from either simulations or data. In our analysis we use the bias parameters from Table 5.3, measured in simulations in [1]. In Eq. (5.40),  $P_{\alpha, \text{shot}}$  is the bispectrum shot noise term. Since one field is the linear field, all contribution to this shot noise comes from the stochastic bias terms in the two galaxy fields  $\delta_g$  in the quadratic estimator, such as  $\varepsilon$  and  $\varepsilon_\delta \delta$  (see App. C.2 or [79] for more discussion about stochastic bias terms). The expression for this shot noise contribution in this case can also be derived from Eq. (C.31) and it takes the form

$$P_{\alpha, \text{shot}}(\vec{k}) = \frac{b_1}{\bar{n}} P_{\text{lin}}(\vec{k}) N_{\alpha\alpha}(\vec{k}) \int_{\vec{q}} \frac{f_\alpha(\vec{q}, \vec{k} - \vec{q})}{2P_{\text{tot}}(\vec{q}) P_{\text{tot}}(\vec{k} - \vec{q})} W(R\vec{q}) W(R(\vec{k} - \vec{q})). \quad (5.41)$$

In Fig. 5.3, we compare theory with simulations for three different values of  $k_{\text{max}}$ . Although for the Fisher analysis in this work, we only use the growth estimator, here we also compare results in simulations for the shift and the tidal estimators. For the growth estimator, we find that the theory predictions agree very well with simulation

results for up to  $k_{\max} = 0.25h \text{ Mpc}^{-1}$  at redshift  $z = 0$ . For the other estimators, we also find reasonably good agreement; however, upon close inspection we can see small disagreements which might arise from higher-order terms ignored in our theory predictions.

Interestingly, for  $k_{\max} = 0.25h \text{ Mpc}^{-1}$ , we can see in Fig. 5.3 that the shape of the cross-correlation of growth estimators with the density field is very similar to the linear power spectrum on large scales. The scale-dependent bias factor in Eq. (5.40) is flat on large scales, indicating that the reconstruction works very well for large  $k_{\max}$ .

### 5.3.3 Auto- and cross-correlations of quadratic estimators

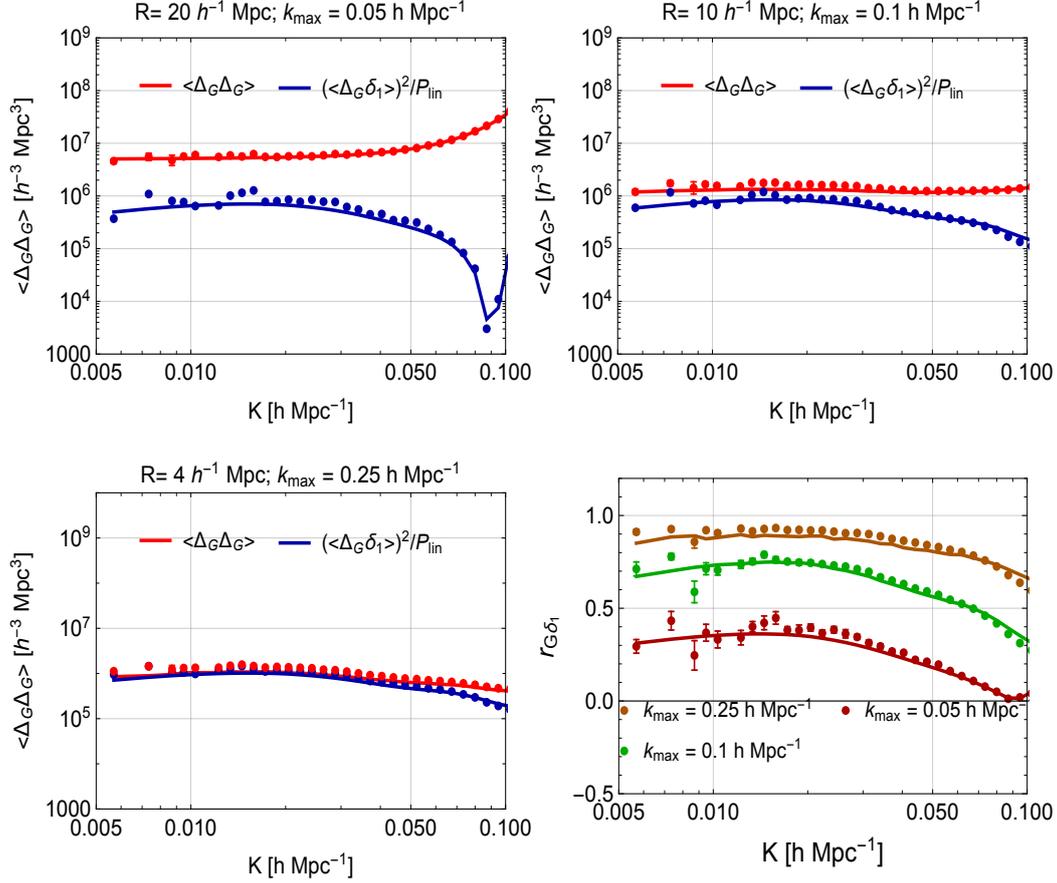
In this section we discuss our results for the auto- and cross-correlations of three quadratic estimators from simulations and compare the results with our linear-order theoretical prediction, given by

$$\begin{aligned} \langle \hat{\Delta}_\alpha(\vec{k}) \hat{\Delta}_\beta(\vec{k}') \rangle' = & \\ b_1^4 N_{\alpha\alpha}(\vec{k}) N_{\beta\beta}(\vec{k}) \int_{\vec{q}} \frac{f_\alpha(\vec{q}, \vec{k} - \vec{q}) f_\beta(\vec{q}, \vec{k} - \vec{q})}{[2P_{\text{tot}}(\vec{q}) P_{\text{tot}}(\vec{k} - \vec{q})]^2} W(R\vec{q})^2 W(R(\vec{k} - \vec{q}))^2 P_{\text{lin}}(\vec{q}) P_{\text{lin}}(\vec{k} - \vec{q}) & \\ + b_1^2 P_{\text{lin}}(\vec{k}) \sum_{i,j} c_i c_j \frac{N_{\alpha\alpha}(\vec{k}) N_{\beta\beta}(\vec{k})}{N_{\alpha i}(\vec{k}) N_{\beta j}(\vec{k})} + P_{\alpha\beta, \text{shot}}(\vec{k}) . & \quad (5.42) \end{aligned}$$

The first term is of order  $\mathcal{O}(\delta_1^4)$ , while the second and third are of order  $\mathcal{O}(\delta_1^6)$ . The third term,  $P_{\alpha\beta, \text{shot}}$ , is the contribution arising from halo shot noise, and is given in App. C.2.

In Fig. 5.4 we compare cross-correlation results from simulations with theory, for the growth, shift, and tidal estimators, using the same three smoothing scales as above. The simulations and theory agree very well up to  $k_{\max} = 0.1h \text{ Mpc}^{-1}$  at  $z = 0$ . For larger  $k_{\max}$  we see good agreement for the growth estimator and reasonable agreement for the tidal and shift estimators. The small disagreement of linear predictions for the tidal and shift estimators with simulations for the higher  $k_{\max}$  show that higher-order terms become important for these estimators. The detailed impact of these higher order corrections from biasing or scale-dependent stochasticity will be the subject of future inquiry. Although we appear to have excellent agreement for the growth term at higher  $k_{\max}$ , to be conservative, we still set the scale  $k_{\max} = 0.1h \text{ Mpc}^{-1}$  at redshift  $z = 0$  in our forecasts in Sec. 5.4. We scale this to other redshifts by making use of

the fact that perturbation theory and the bias expansion at a given order will be valid at higher  $k$  for higher redshifts.



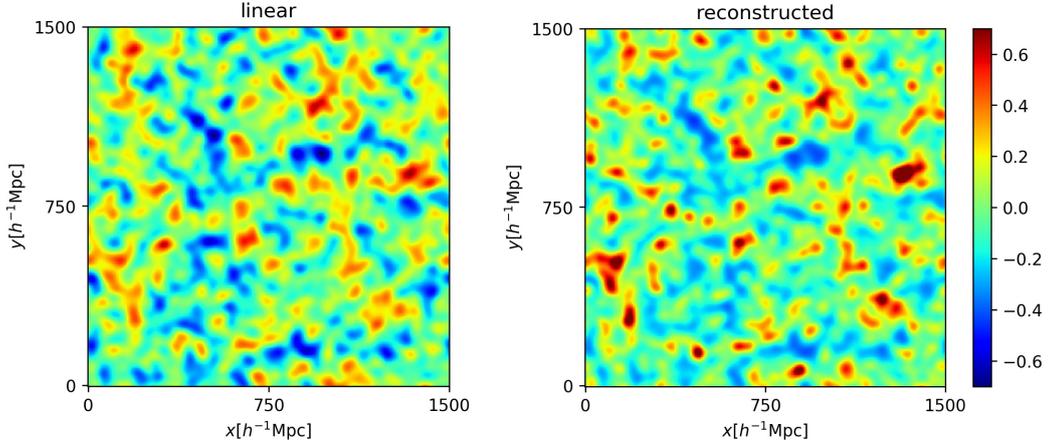
*Fig. 5.5* Comparison of the auto power spectrum of the growth estimator  $\hat{\Delta}_G$ , normalised by  $N_{GG}$  computed from theory (in red), with  $(\langle \hat{\Delta}_G \delta_1 \rangle)^2 / P_{\text{lin}}$  (in blue). We compare simulation results (points) with theory predictions (lines) for the same smoothing scales as Figs. 5.3 and 5.4. We again find excellent agreement between simulations and theory. In the bottom right panel, we plot the cross-correlation coefficient  $r_{G\delta_1}$  between the growth estimator and the linear density field for three smoothing scales. We see that  $r_{G\delta_1} > 0.9$  for  $R = 4 h^{-1} \text{ Mpc}$  which is why in the bottom-left panel,  $\langle \hat{\Delta}_G \hat{\Delta}_G \rangle$  is signal dominated.

In Fig. 5.5, we plot the auto spectra of the growth estimator, normalized with  $N_{GG}$  (unlike in the previous plots), in order to compare them to an approximation of the signal power spectrum, given by the second term in Eq. (5.42) (the first and third terms represent noise). Since the contribution of the cross-shot noise is small, the signal part can be approximated by cross-correlating the growth estimator with the linear density field and dividing it by the linear power spectrum to ensure the correct normalization, i.e.  $(\langle \hat{\Delta}_G \delta_1 \rangle)^2 / P_{\text{lin}}$ ; we show this in blue in Fig. 5.5. For the

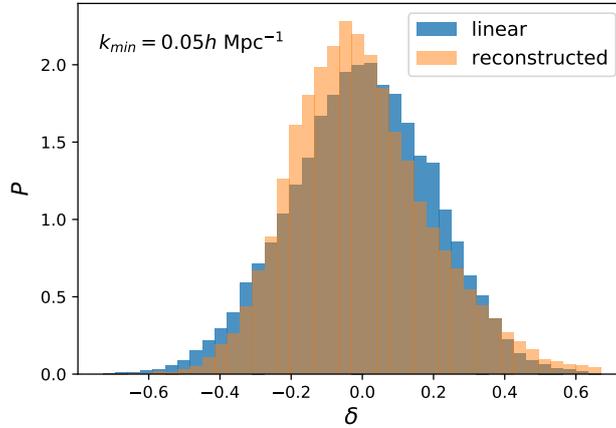
two larger smoothing scales, the spectra of the estimator are dominated by the noise contribution (which is white at low  $k$ ). The excellent agreement between theory (red solid lines) and simulations (red points) for all smoothing scales serves as an additional verification that the reconstruction procedure is working as expected for reasonable values of  $k_{\max}$ . In addition to the auto spectra, to check how well the reconstruction is working, we plot the cross-correlation coefficients between the growth estimator and the linear density field in the bottom-right panel of Fig. 5.5 for three different  $k_{\max}$ . The cross-correlation coefficient for low  $k_{\max}$  is very low,  $r_{G\delta_1} < 0.4$ , explaining why the auto spectra in the top-left panel are noise dominated. However, for the highest  $k_{\max}$  we consider,  $0.25h \text{ Mpc}^{-1}$ , the cross-correlation coefficient is  $r_{G\delta_1} > 0.9$ , which explains why the reconstruction works very well and the auto spectra for high  $k_{\max}$  are signal dominated.

### 5.3.4 Visualization of reconstructed field

To visualize how well we are reconstructing the linear density field on large-scales in simulations, we compare 2D slices of thickness  $6h^{-1}\text{Mpc}$  of the linear density field and the reconstructed field in Fig. 5.6. We perform the reconstruction using  $k_{\max} = 0.25h \text{ Mpc}^{-1}$ , i.e., smoothing at a scale of  $R = 4h^{-1} \text{ Mpc}$ . In the visualization, we apply an external smoothing of  $R = 20h^{-1} \text{ Mpc}$  to both the linear field and the reconstructed field, which removes all modes with  $k > 0.05h \text{ Mpc}^{-1}$ . Our comparison of the linear and reconstructed fields in Fig. 5.6 shows that the reconstruction indeed recovers most of the large-scale features in the linear density field. In Fig. 5.7 we show histograms, probing the one-point probability distribution functions, of the linear density field and the reconstructed field. We see that the reconstructed field is nearly Gaussian, partially justifying our approximation of a Gaussian likelihood in the next section.



*Fig. 5.6* 2D slices of the 3D linear density field (left panel) and the growth estimator  $\hat{\Delta}_G$  (right panel). For the growth estimator we used  $R = 4h^{-1}$  Mpc smoothing which corresponds to  $k_{\max} = 0.25h$  Mpc $^{-1}$ . We apply an external smoothing of  $R = 20h^{-1}$  Mpc to both the linear and reconstructed fields. As expected, we find that the reconstruction reproduces many of the large-scale features in the linear density field.



*Fig. 5.7* Probability distribution functions (histograms) of the linear density field and the reconstructed field from the halo density field of mass bin I. As in Fig. 5.6, we use  $k_{\max} = 0.25h$  Mpc $^{-1}$  for the reconstruction and apply an external smoothing scale of  $R = 20h^{-1}$  Mpc to both the linear field and the reconstructed field. The PDFs of the reconstructed field are scaled to have the same variance as the linear field, and shifted to have mean 0. We find that the PDF of the reconstructed field is very close to Gaussian. Note that here we have applied a low- $k$  cutoff to the modes used for reconstruction of  $k_{\min} = 0.05h$  Mpc $^{-1}$  in order to match the approach in our forecast section below.

## 5.4 Forecasts

### 5.4.1 Fisher matrix setup

To perform a Fisher forecast, we make the usual assumption that the measured tracer overdensity  $\delta_g$  and the reconstructed field  $\delta_r$  obey a Gaussian likelihood. For the matter and galaxy field, this approximation is partially justified by the fact that we are analyzing very large scales; for the reconstruction noise, this is partially justified by the fact that the reconstruction sums over a large number of mode pairs, so that to some extent the central limit theorem applies (although the pairs may not all be independent). Figure 5.7 supplies additional evidence that a simple Fisher forecast is sufficient, in that the PDFs of the density field (smoothed to correspond with our analysis range) do not greatly deviate from a Gaussian. This indicates that the influence of higher moments of the density field and noise is comparatively small for the purposes of a forecast.

Making this approximation and including the fact that  $\delta_g$  has zero statistical mean, the Fisher matrix per mode  $\vec{K}$  and redshift  $z$  is given by (e.g. [273])

$$\tilde{F}_{ab}(\vec{K}, z) = \frac{1}{2} \text{Tr} \left[ \partial_a C(\vec{K}, z) C^{-1}(\vec{K}, z) \partial_b C(\vec{K}, z) C^{-1}(\vec{K}, z) \right], \quad (5.43)$$

where  $C$  is the total (signal plus noise) covariance matrix for our data vector  $\vec{d}(\vec{K}) = (\delta_g(\vec{K}), \delta_r(\vec{K}))^T$ ,  $\text{Tr}$  is the trace matrix operator,  $\partial_b C(\vec{K}, z) \equiv \frac{\partial}{\partial b} C(\vec{K}, z)$ , and  $a, b$  are the parameters on which our quantities depend (in this case,  $f_{\text{NL}}$  and bias parameters). If the data vector is drawn from a Gaussian distribution and nothing is known about the parameters, then the inverse of the Fisher matrix gives the covariance matrix of the parameters, and the square root of the diagonal elements of  $F^{-1}$  give the errorbars on the parameters and represent the minimum error achievable. Our goal is to calculate this minimum error, as it will determine our best ability to constrain parameters.

In reality, we do not just measure a single mode, but we measure several modes whose information can be combined together in an integrated Fisher matrix for a specific redshift bin, i.e.

$$F_{ab}(z) = \frac{V}{(2\pi)^2} \int_{K_{\min}}^{K_{\max}} dK \int_{-1}^1 d\mu K^2 \tilde{F}_{ab}(K, \mu, z). \quad (5.44)$$

Here  $V$  is the survey volume,  $K_{\min}$  and  $K_{\max}$  are the minimum and maximum moduli of the modes probed, and we already integrated over the azimuthal direction, supposing no dependence from it in the integrand.

For our specific case, the original field  $\delta_g$  and the reconstructed field  $\delta_r$  will give the total covariance matrix (which only depends on the magnitude of  $\vec{K}$ )

$$C(K, z) = \begin{bmatrix} C^{\text{gg}}(K, z) & C^{\text{gr}}(K, z) \\ C^{\text{gr}}(K, z) & C^{\text{rr}}(K, z) \end{bmatrix}, \quad (5.45)$$

with elements

$$C^{\text{gg}}(K, z) = \left( b_1(z) + \frac{c_{01} f_{\text{NL}}}{M(K, z)} D(z) \right)^2 P_{\text{lin}}(K, z) + P_{\text{gg,shot}}(K, z), \quad (5.46)$$

$$C^{\text{gr}}(K, z) = \left( b_1(z) + \frac{c_{01} f_{\text{NL}}}{M(K, z)} D(z) \right) b_r(K, z) P_{\text{lin}}(K, z) + P_{\text{gr,shot}}(K, z), \quad (5.47)$$

$$C^{\text{rr}}(K, z) = b_r(K, z)^2 P_{\text{lin}}(K, z) + N_{\text{GG}}(K, z) + P_{\text{rr,shot}}(K, z), \quad (5.48)$$

where

$$b_r \equiv b_1 \left( c_G + \sum_{\beta \neq G} c_\beta \frac{N_{\text{GG}}}{N_{\text{G}\beta}} \right), \quad (5.49)$$

and the sum runs over the mode-couplings found in Table 5.2. We do not include redshift space distortions in these expressions; see Sec. 5.4.3 for discussion.

The tracer shot noise is simply

$$P_{\text{gg,shot}}(K, z) = \frac{1}{\bar{n}(z)}, \quad (5.50)$$

where  $\bar{n}$  is the comoving number density of observed tracers, while  $P_{\text{rr,shot}}$  and  $P_{\text{gr,shot}}$  are given in Eqs. (C.26)–(C.27) and (C.31)–(C.32) respectively. We will neglect the dependence of the reconstruction shot noise on  $f_{\text{NL}}$ . This is because in general these shot noise terms include the small-scale tracer power spectrum, whose response to a change of  $f_{\text{NL}}$  is negligible compared to the response experienced by the large-scale power spectrum. Moreover, even when the large-scale tracer power spectrum enters the reconstruction shot noise, as in  $P_{\text{rr,shot}}$  where there is a coupling between large and small scales as we explain in Appendix C.2, a small change from  $f_{\text{NL}} = 0$ , our fiducial value, is barely detectable. In principle, it may be possible extract additional information from the  $f_{\text{NL}}$ -dependence of the shot noise contributions, but this will likely be difficult in practice, and therefore we conservatively choose not to consider these contributions as observables.

Substituting Eq. (5.45) into Eq. (5.43), we can derive an explicit formula for the Fisher matrix per mode for our case, which can then be inserted into Eq. (5.44):

$$\begin{aligned} \tilde{F}_{ab} = & \frac{1}{2} \left( \frac{1}{C^{rr}C^{gg}(1-r_{cc}^2)} \right)^2 \\ & \times \left[ C^{gg} \left\{ \partial_b C^{gr} \left( -C^{gr} \partial_a C^{rr} + C^{rr} \partial_a C^{gr} \right) + \partial_b C^{rr} \left( C^{gg} \partial_a C^{rr} - C^{gr} \partial_a C^{gr} \right) \right\} \right. \\ & - C^{gr} \left\{ \partial_b C^{gg} \left( -C^{gr} \partial_a C^{rr} + C^{rr} \partial_a C^{gr} \right) + \partial_b C^{gr} \left( C^{gg} \partial_a C^{rr} - C^{gr} \partial_a C^{gr} \right) \right\} \\ & - C^{gr} \left\{ \partial_b C^{gr} \left( -C^{gr} \partial_a C^{gr} + C^{rr} \partial_a C^{gg} \right) + \partial_b C^{rr} \left( C^{gg} \partial_a C^{gr} - C^{gr} \partial_a C^{gg} \right) \right\} \\ & \left. + C^{rr} \left\{ \partial_b C^{gg} \left( -C^{gr} \partial_a C^{gr} + C^{rr} \partial_a C^{gg} \right) + \partial_b C^{gr} \left( C^{gg} \partial_a C^{gr} - C^{gr} \partial_a C^{gg} \right) \right\} \right], \quad (5.51) \end{aligned}$$

where  $r_{cc}$  is the g-r cross-correlation coefficient:

$$r_{cc} \equiv \frac{C^{gr}}{\sqrt{C^{gg}C^{rr}}} . \quad (5.52)$$

For  $a = b$ , we obtain

$$\begin{aligned} \tilde{F}_{aa} = & \frac{1}{2(1-r_{cc}^2)^2} \left[ \left( \frac{\partial_a C^{gg}}{C^{gg}} - 2r_{cc}^2 \frac{\partial_a C^{gr}}{C^{gr}} \right)^2 + 2r_{cc}^2 (1-r_{cc}^2) \left( \frac{\partial_a C^{gr}}{C^{gr}} \right)^2 \right. \\ & \left. + 2r_{cc}^2 \frac{\partial_a C^{rr}}{C^{rr}} \left( \frac{\partial_a C^{gg}}{C^{gg}} - 2 \frac{\partial_a C^{gr}}{C^{gr}} \right) + \left( \frac{\partial_a C^{rr}}{C^{rr}} \right)^2 \right] . \quad (5.53) \end{aligned}$$

On the other hand, if we only use  $\delta_g$ , we get

$$\tilde{F}_{aa}^{(g \text{ only})} = \frac{1}{2} \left( \frac{\partial_a C^{gg}}{C^{gg}} \right)^2 . \quad (5.54)$$

### 5.4.2 Analytical derivation of cosmic variance cancellation

Cosmic variance cancellation will occur in the limit of low noise on the measured fields – that is, low reconstruction noise on the quadratic estimator, and low galaxy shot noise. To investigate this case analytically, let us work in the limit of very low shot noise, so that

$$\begin{aligned} C^{gg}(K) &= b_g(K)^2 P_{\text{lin}}(K) , \\ C^{gr}(K) &= b_g(K) b_r(K) P_{\text{lin}}(K) , \\ C^{rr}(K) &= b_r(K)^2 P_{\text{lin}}(K) + N_{\text{GG}}(K) . \end{aligned} \quad (5.55)$$

Further, let us assume that  $f_{\text{NL}}$  is the only unknown parameter. If we define

$$x(K) \equiv \frac{N_{\text{GG}}(K)}{b_{\text{r}}(K)^2 P_{\text{lin}}(K)}, \quad R_p(K) \equiv \left( \frac{\partial_{f_{\text{NL}}} b_{\text{r}}(K)}{b_{\text{r}}(K)} \right) \left( \frac{\partial_{f_{\text{NL}}} b_{\text{g}}(K)}{b_{\text{g}}(K)} \right)^{-1}, \quad (5.56)$$

where  $x(K)$  is the inverse of the signal-to-noise per mode of the reconstructed field and  $R_p(K)$  is a measure of similarity between the response of the bias of the reconstructed field and the one of the original tracer field, then a short calculation gives the unmarginalized errorbar on  $f_{\text{NL}}$  per  $K$ -mode:

$$\sigma_{f_{\text{NL}}}^2(K) = \sigma_{f_{\text{NL}}, \text{g only}}^2(K) \frac{2x(K)}{(R_p(K) - 1)^2} \frac{1}{1 + 2(R_p(K) - 1)^{-2} x(K)}, \quad (5.57)$$

where  $\sigma_{f_{\text{NL}}, \text{g only}} = [F_{\text{aa}}^{(\text{g only})}]^{-1/2}$ .

Let us investigate the general behavior of this equation in some limiting cases. If  $(R_p - 1)^{-2} x$  is small, the  $R_p < 0$  case (when  $\partial_{f_{\text{NL}}} b_{\text{r}}$  and  $b_{\text{r}}$  have opposite signs) will result in smaller errorbars than the  $R_p > 0$  case, because the signatures of  $f_{\text{NL}}$  in  $b_{\text{r}}$  and  $b_{\text{g}}$  will be more distinguishable in that case. Expanding Eq. (5.57) in the limit of small  $(R_p - 1)^{-2} x$  gives

$$\sigma_{f_{\text{NL}}}^2(K) = \sigma_{f_{\text{NL}}, \text{g only}}^2(K) \frac{2x(K)}{(R_p(K) - 1)^2} \sum_{n=0}^{\infty} \left[ -2(R_p(K) - 1)^{-2} x(K) \right]^n. \quad (5.58)$$

As we show in Appendix C.4,  $N_{\text{GG}} \propto k_{\text{max}}^{-3}$  in the low- $K$  limit, so that we arrive at

$$\lim_{x \rightarrow 0} \sigma_{f_{\text{NL}}}^2(K) \propto 2\sigma_{f_{\text{NL}}, \text{g only}}^2(K) \left[ k_{\text{max}}^{-3} + \mathcal{O}(k_{\text{max}}^{-6}) \right], \quad (5.59)$$

where we assume that  $R_p - 1$  varies slowly with  $K$ . This demonstrates that constraints on  $f_{\text{NL}}$  that use both reconstructed modes and modes of the original tracer will improve on a tracer-only analysis in a way that is only limited by the noise on the reconstructed modes (if shot noise is negligible).

Cosmic variance cancellation clearly requires that  $\delta_{\text{r}}$  and  $\delta_{\text{g}}$  are measured at the same wavenumber and in the same volume. To verify this, we can repeat the derivation above with  $C^{\text{gr}} = 0$ , corresponding to  $\delta_{\text{r}}$  and  $\delta_{\text{g}}$  being measured in different volumes. In this case, Eq. (5.57) becomes

$$\sigma_{f_{\text{NL}}}^2(K) = \sigma_{f_{\text{NL}}, \text{g-only}}^2(K) \frac{[1 + x(K)]^2}{R_p(K)^2 + [1 + x(K)]^2}, \quad (5.60)$$

which approaches a finite limit as  $x \rightarrow 0$ ; thus, the improvement realized in Eq. (5.59) is only possible if  $\delta_{\text{r}}$  and  $\delta_{\text{g}}$  can be compared mode-by-mode in the same volume.

	DESI-like	MegaMapper-like		PUMA-like	
	$0.6 < z < 1.6$	$2 < z < 2.5$	$4.5 < z < 5$	$2 < z < 3$	$5 < z < 6$
<b>Survey parameters</b>					
Survey volume ( $\text{Gpc}^3$ )	100	80	66	266	203
Mean galaxy density $\bar{n}$ ( $\text{Mpc}^{-3}$ )	$10^{-4}$	$6 \times 10^{-4}$	$2 \times 10^{-5}$	$2 \times 10^{-3}$ ( $6 \times 10^{-3}$ )	$1 \times 10^{-3}$ ( $2 \times 10^{-2}$ )
$K_{\text{max}}$ for $f_{\text{NL}}$ constraint ( $h \text{Mpc}^{-1}$ )	0.05	0.08	0.14	0.09	0.15
$k_{\text{max}}$ for reconstruction ( $h \text{Mpc}^{-1}$ )	0.15	0.24	0.4	0.26	0.47
<b>Fiducial bias parameters</b>					
$b_1$	1.6	2.9	7.0	2.1	3.7
$b_2$	-0.30	1.1	17	0.041	2.8
$b_{s^2}$	-0.17	-0.54	-1.7	-0.31	-0.77
$b_{11}^E$	-3.0	-2.5	37	-3.5	0.58
$b_{02}^E$	-14	-21	85	-19	-16

*Table 5.4* Survey characteristics used for our main forecasts. The DESI-like survey is based on the expected DESI emission-line galaxy sample, the MegaMapper-like survey is a next-generation survey targeting high-redshift “dropout” galaxies, and the PUMA-like survey represents a future 21 cm intensity mapping effort over half the sky. We marginalize over  $b_1$ ,  $b_2$ , and  $b_{s^2}$  in our forecasts, and determine  $b_{11}^E$  and  $b_{02}^E$  using the relationships in Sec. 5.2.2. For the PUMA-like forecast, the main  $\bar{n}$  values represent effective number densities that reproduce the same noise level as the sum of shot and instrumental noise power at  $k = k_{\text{max}}$ , while the expected physical number densities are shown in parentheses. For this forecast, we also consider the effects of the so-called “foreground wedge” that will prevent direct measurement of certain modes. See main text for details.

### 5.4.3 Assumptions and experimental configurations

#### Scales

In each forecast, for measuring  $f_{\text{NL}}$ , we use  $\delta_{\text{g}}$  modes and reconstructed modes with wavenumber  $K$  satisfying  $K_{\text{min}} < K < K_{\text{max}}$ , and we also use reconstructed modes with  $K_{\text{f}} < K < K_{\text{min}}$ , where  $K_{\text{f}} \approx 0.002h \text{Mpc}^{-1}$  is the lowest measurable wavenumber within each survey volume. In this way,  $K_{\text{min}}$  accounts for possible systematic effects that can prevent direct measurements of  $\delta_{\text{g}}$  on large scales, but that do not impede reconstruction of these large-scale modes using smaller-scale correlations; an example is foreground contamination for intensity mapping experiments, which has been a primary motivator for other work on reconstruction methods [292, 293, 91, 155, 141, 188]. As input to the density-field reconstruction, we use modes with wavenumber  $k$  satisfying  $K_{\text{max}} < k < k_{\text{max}}$ . We consider a range of possible  $K_{\text{min}}$  values in our forecasts, while  $k_{\text{max}}$  and  $K_{\text{max}}$  are fixed for each survey, as described below.

#### Surveys

In our main forecasts, we consider three galaxy surveys, with properties summarized in Table 5.4. The first is similar to the emission-line galaxy sample expected from DESI [7]. For this survey, following [192], we consider  $14000 \text{deg}^2$  of sky area over

$0.6 < z < 1.6$ , which translates into a total comoving volume of roughly  $100 \text{ Gpc}^3$  and a mean redshift of  $\bar{z} \approx 1$ . We use a mean galaxy number density of  $\bar{n} = 10^{-4} \text{ Mpc}^{-3}$ , obtained by dividing the expected total number of redshifts in the DESI ELG sample ( $1.7 \times 10^7$ , from [7]) by the survey volume, and assume a mean linear galaxy bias of  $b_1 = 1.6$ . We take  $K_{\text{max}} = 0.05h \text{ Mpc}^{-1}$ , since linear bias is expected to be an acceptable approximation for  $K < K_{\text{max}}$  at  $z = 1$ , and  $k_{\text{max}} = 0.15h \text{ Mpc}^{-1}$ , since our quadratic bias expansion is valid for  $k < k_{\text{max}}$  at  $z = 1$  (see Sec. 5.3 for justification based on simulations).

The second survey, which we call ‘‘MegaMapper-like’’, is modelled on proposals for a next-generation spectroscopic survey targeting high-redshift ‘‘dropout’’ galaxies in the southern hemisphere [288, 90, 232]. For this, we assume a  $14000 \text{ deg}^2$  survey, and separately consider two redshift bins, at  $2 < z < 2.5$  and  $4.5 < z < 5$ , which have volumes of  $80 \text{ Gpc}^3$  and  $66 \text{ Gpc}^3$  respectively. The mean number density and linear bias in each bin are obtained from averages of the values at the bin edges, taken from Table 1 of [90]; this yields  $\bar{n} = 6 \times 10^{-4} \text{ Mpc}^{-3}$  and  $b_1 = 2.9$  for the lower-redshift bin, and  $\bar{n} = 2 \times 10^{-5} \text{ Mpc}^{-3}$  and  $b_1 = 7.0$  for the higher-redshift bin. For  $K_{\text{max}}$  and  $k_{\text{max}}$ , we scale the DESI values using the ratio of linear growth factors between the mean redshifts of each redshift bin, to account for the increased range of validity of our perturbative expressions at higher redshift.<sup>11</sup>

The third survey is based on specifications for PUMA, an envisioned radio interferometer designed for 21 cm intensity mapping [19, 261]. We assume a survey over half the sky, and again consider two redshift bins, this time at  $2 < z < 3$  and  $5 < z < 6$ , with volumes  $266 \text{ Gpc}^3$  and  $203 \text{ Gpc}^3$  respectively. For simplicity, we treat this survey as observing galaxy positions directly, rather than brightness temperature (which is just a rescaled biased tracer of the matter density). To do so, we set the noise contribution to the tracer power spectrum  $P_{\text{gg}}$  to equal the sum of the shot noise and instrumental noise power spectra computed using the PUMA noise calculator<sup>12</sup>, evaluated at  $k = k_{\text{max}}$  in each redshift bin. We quote an effective number density that would result in the same noise level in Table 5.4. When computing the shot noise contributions to  $P_{\text{gr}}$  and  $P_{\text{rr}}$ , we use the expected number densities of 21 cm emitters, also taken from the PUMA noise calculator and quoted in parentheses in Table 5.4. For the linear bias in each bin, we use values from Fig. 33 of [19], evaluated at the

<sup>11</sup>The argument for this  $k_{\text{max}}$  scaling is as follows. Nonlinear displacements are one of the physical effects that limit the accuracy of perturbation theory, and it is natural to assume that the redshift-dependence of  $k_{\text{max}}$  will scale with the rms displacement, which we compute in the Zel’dovich approximation. Under these assumptions,  $k_{\text{max}}(z) \propto D^{-1}(z)$ . In reality, the scaling of  $k_{\text{max}}$  with redshift is more complicated, involving the power spectrum tilt at the relevant wavenumbers (e.g. [49]), but the simple growth factor scaling we use here should at least be roughly indicative of the useful scales for our forecasts.

<sup>12</sup><https://github.com/slosar/PUMANoise>

mean redshifts. As for the MegaMapper-like survey, we scale  $K_{\max}$  and  $k_{\max}$  from DESI by the appropriate ratios of linear growth factors.

One could also consider applying our reconstruction procedure to data from the SPHEREx satellite [82], which aims to map large-scale structure by using low-resolution spectroscopy to determine the redshifts of over 500 million galaxies. The total SPHEREx galaxy sample will have higher number density than DESI ELGs by at least a factor of 5, but many of these galaxies will have significant redshift uncertainties that will suppress the measured clustering power in a scale-dependent way, and this would need to be included in our formalism. On the other hand, angular variations in the mean redshift uncertainty may also imprint spurious clustering signals on the data, and long-mode reconstruction is a promising avenue to probe or mitigate this. We intend to explore these issues in future work.

In our derivation of stochastic contributions to the noise of the estimator and the cross-correlation between estimator and galaxy fields in App. C.2, we assume that the noise is Poissonian, i.e., that  $\langle \varepsilon \varepsilon \rangle = (2\pi)^3 \delta_{\mathbf{D}}(\vec{K} + \vec{K}')/\bar{n}$ . There is evidence for halo stochasticity being sub-Poissonian for high-mass haloes and super-Poissonian for low mass haloes [106, 25]. Since the stochasticity corrections arise from small-scale exclusion and higher-order biases, the actual shot noise levels cannot be theoretically predicted, implying that it may be advisable to marginalize over the stochasticity parameter(s). This approach is indeed adopted by some for the  $f_{\text{NL}}$  forecasting literature (e.g. [51]) but certainly not all of it (e.g. [237, 192]). Here we decide to fix the stochasticity parameters to their fiducial Poissonian values and defer a more detailed investigation of the impact of noise corrections on the reconstructed fields to future work. We do note however, that we expect the impact of shot noise marginalization to be rather small, since we do not include the additional non-Gaussian signal arising in combination with stochastic terms in Eqs. (5.46-5.48).

## 21 cm foregrounds

An additional consideration for 21 cm intensity mapping is the presence of foreground radiation, predominantly synchrotron from our own galaxy, that is brighter than the cosmological signal by several orders of magnitude. These foregrounds are extremely smooth in frequency, which implies that they mainly populate Fourier modes with low line-of-sight wavenumber  $k_{\parallel}$ ; these modes will therefore likely not be usable for cosmology. Furthermore, the chromatic properties of interferometers generically spread foreground power from the low- $k_{\parallel}$  modes into a wedge-shaped region in the  $k_{\parallel} - k_{\perp}$  plane (e.g. [204, 161, 162]), although this contamination can be removed with sufficiently precise instrumental calibration (e.g. [251, 94]).

For constraining  $f_{\text{NL}}$ , the wedge will have two effects: it will reduce the number of short-wavelength modes available for the quadratic estimator, therefore increasing the noise  $N_{\text{GG}}$  on the reconstructed modes, and it will also reduce the number of long-wavelength  $\delta_{\text{g}}$  modes available for measuring the scale-dependent bias induced by primordial non-Gaussianity.

We account for both effects in our forecasts for the PUMA-like survey, assuming a foreground wedge defined by 3 times the primary beam width, following [19]; see Appendix C.5.1 for details of how this is implemented in our computations. In addition, we perform forecasts that ignore the wedge, to represent the case when it can be completely eliminated via calibration. We account for lost low- $K_{\parallel}$  modes in two ways: either by restricting  $\delta_{\text{g}}$  to have  $K > K_{\parallel, \text{min}}$ , or by approximating  $K_{\parallel, \text{min}}$  as an isotropic  $K_{\text{min}}$ , matching our procedure for DESI and MegaMapper. The former approach is more realistic, while the latter is easier to compare with the other surveys, so we present the latter in the main text, and the former in Appendix C.5.2.

### Bias parameters

For every survey, to perform forecasts, we assume a fiducial value of the quadratic bias parameter  $b_2$  derived from the fitting formula of [148], which was fit to halo bias in separate-universe simulations over the range  $1 \lesssim b_1 \lesssim 10$ :

$$b_2(b_1) = 2 \left( 0.412 - 2.143b_1 + 0.929b_1^2 + 0.008b_1^3 \right) , \quad (5.61)$$

where the extra factor of 2 arises from our different definition of  $b_2$  compared to [148]. The fiducial value of the tidal bias  $b_{s^2}$  is found from

$$b_{s^2} = -\frac{2}{7} (b_1 - 1) , \quad (5.62)$$

which assumes that the tidal bias in Lagrangian space is zero. In our forecasts,  $b_1$ ,  $b_2$ , and  $b_{s^2}$  are allowed to vary independently (i.e. are marginalized over when we estimate uncertainties on  $f_{\text{NL}}$ ), while  $b_{11}$  and  $b_{02}$  are assumed to obey the relationships in Eqs. (5.32)-(5.33) and (5.34)-(5.36). We take wide, flat priors on  $b_1$ ,  $b_2$ , and  $b_{s^2}$ ; we have also implemented 10% Gaussian priors on  $b_2$  and  $b_{s^2}$ , but these have a negligible effect on our baseline results.

### Redshift space distortions

The line-of-sight component of a galaxy's position is observationally inferred from the galaxy's redshift, and the associated "redshift-space distortions" of  $\delta_{\text{g}}$  should be

included in a full treatment of the observed galaxy clustering. The leading-order effect is to add a  $f\mu^2$  term to the linear bias of  $\delta_g$ , such that Eq. (5.37) is modified to

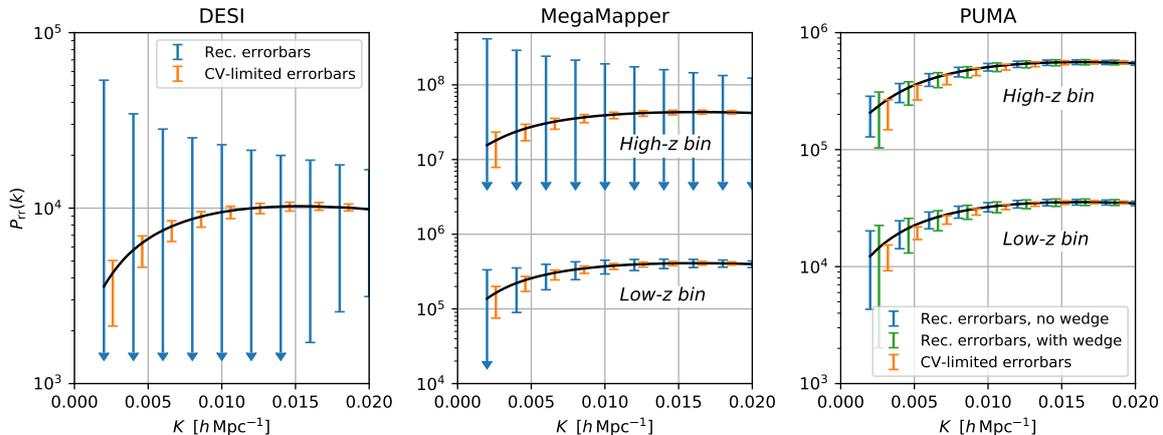
$$\delta_g(\vec{k}) = \left[ b_1 + f_{\text{NL}} \frac{c_{01}}{M(k)} + f\mu^2 \right] \delta_1(\vec{k}) + \dots, \quad (5.63)$$

where  $f \equiv d \log D / d \log a$ ,  $\mu \equiv k_{\parallel} / k$ , and  $D$  is the linear growth factor [138]. Higher-order effects will create additional mode-couplings that can be described in perturbation theory (e.g. [207, 37]). In a real tracer catalogue, there will also be line-of-sight-dependent selection effects that can be treated perturbatively [80].

We do not include any of these effects in our baseline forecasts, leaving them for future work. However, as a first step in this direction, we have checked the impact of including the Kaiser term. This raises the reconstruction noise  $N_{\text{GG}}$  by increasing  $P_{\text{gg,tot}}$  in the denominator of Eq. (5.17), while also increasing the amplitude of  $P_{\text{gg}}$  and  $P_{\text{gr}}$ , thereby increasing the signal to noise on those quantities. For all surveys we consider, the former effect overcomes the latter, with the result that  $\sigma(f_{\text{NL}})$  increases by roughly 10%, and the improvement in  $\sigma(f_{\text{NL}})$  from including reconstructed modes decreases by no more than the same amount. Additional mode-couplings from nonlinear redshift-space effects will likely dominate over this change, and a detailed analysis will be worthwhile to pursue, especially since some of these mode-couplings could potentially carry additional information about  $f_{\text{NL}}$  [51].

#### 5.4.4 Expected precision on reconstructed modes

Aside from primordial local non-Gaussianity, there are many other applications of reconstructing large-scale modes, including more general constraints on cosmology, tests of predictions for the power spectrum on the largest scales, calibration of photometric redshifts [188], cross-correlations with other tracers (such as kSZ fluctuations in the CMB, e.g. [155]), and removing contamination from measurements of lensing of 21 cm fluctuations [91]. To represent the general utility of reconstructed modes from different surveys, in Fig. 5.8 we show the expected precision on the auto power spectrum of the reconstructed modes (plotted using the fiducial bias parameters from Table 5.4), computed in wavenumber bins with  $\Delta K = 0.002h \text{ Mpc}^{-1}$ . While these errorbars are substantial for  $K \lesssim 0.01h \text{ Mpc}^{-1}$  in DESI and the high- $z$  bin of MegaMapper, the precision is expected to be much better for MegaMapper at low  $z$  and across the entire redshift range of PUMA, with most errorbars approaching the cosmic variance limit. This will enhance many scientific applications of these surveys, particularly for PUMA, where large-scale modes can be reconstructed at high precision even in the presence of the foreground wedge.

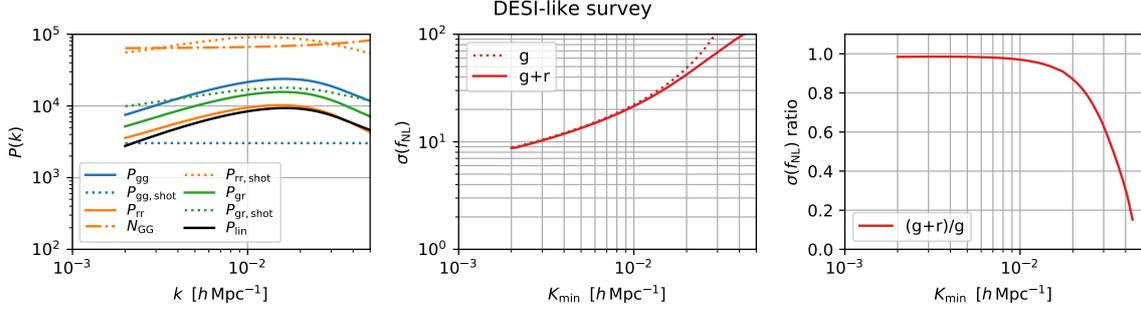


*Fig. 5.8* Expected errorbars on the reconstructed power spectrum  $P_{rr}$  for the surveys and redshift bins we consider (*blue*), along with cosmic-variance-limited errorbars (*orange*), computed for bandpowers with  $\Delta K = 0.002h \text{ Mpc}^{-1}$ . Downward arrows indicate errorbars whose lower limits fall outside of the  $y$  axis range. High-precision measurements of the power spectrum of reconstructed modes will be possible in several cases, even in the presence of a 21 cm foreground wedge for PUMA.

Reference [19] also estimates the total signal to noise in reconstructed modes from PUMA over  $1 < z < 6$ , following the methodology of [91], finding  $\mathcal{O}(1300)$  in the no-wedge case and  $\mathcal{O}(500)$  for the same wedge model we use here. For comparison, we find a total S/N of 135 (108) for  $2 < z < 3$  and 161 (134) for  $5 < z < 6$  in the no-wedge (wedge) case. A direct comparison between the two sets of forecasts is difficult, because they use several distinct approximations: [19] treats the 21 cm brightness temperature as a linearly biased tracer of the matter density, while we have incorporated second-order biasing; [19] neglects the shot noise contribution to the reconstructed mode power spectrum, while we include it; [19] bias-harden their results against mode-couplings from gravitational lensing, while we do not; and, most importantly, [19] only consider reconstruction of modes that are purely transverse to the line of sight ( $k_{\parallel} = 0$ ), while we use a 3D reconstruction formalism. Nevertheless, both forecasts reach the same broad conclusion that PUMA will be able to reconstruct long-wavelength density modes with total signal to noise of several hundred, which is strong motivation for continued studies of the density reconstruction method we have presented in this chapter.

### 5.4.5 Results: constraints on non-Gaussianity

#### DESI

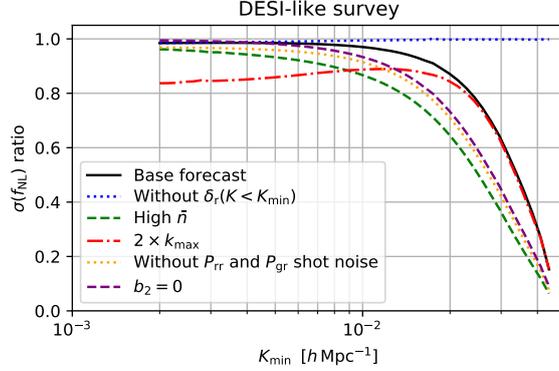


*Fig. 5.9* Forecasts for a DESI-like survey. *Left:* Signal and noise power spectra involved in the forecast. The galaxy auto spectrum is well above the shot noise, while the auto spectrum of reconstructed modes ( $P_{rr}$ ) is roughly an order of magnitude below both the reconstruction noise ( $N_{GG}$ ) in the quadratic estimator and the shot noise contribution to the estimator variance. *Center:* Expected constraints on  $f_{NL}$  when only  $\delta_g$  is used (*solid*), or when  $\delta_r$  is also used (*dotted*). We assume that  $\delta_g(\vec{K})$  cannot be directly measured for  $K < K_{min}$ , and marginalize over the  $b_1$ ,  $b_2$ , and  $b_{s,2}$  bias parameters. *Right:* Ratio of  $\delta_g + \delta_r$  and  $\delta_g$ -only cases from the center panel. We only notice an improvement for higher values of  $K_{min}$ , corresponding to using  $\delta_r$  but not  $\delta_g$  at  $K < K_{min}$ .

Figure 5.9 shows the results of our forecasts for the DESI-like survey. The left panel shows the various power spectra of interest, of linear matter density, galaxy number density, and reconstructed matter density modes, along with the cross spectrum between galaxies and reconstructed modes. This panel also shows the shot noise on  $P_{gg}$ ,  $P_{gr}$ , and  $P_{rr}$ , as well as the statistical noise ( $N_{GG}$ ) on reconstructed modes. For DESI, the galaxy power spectrum is well above the shot noise, while the reconstructed power spectrum is about an order of magnitude lower than the reconstruction noise. Despite the fact that galaxy shot noise is below  $P_{gg}$ , the shot noise contributions for both  $P_{gr}$ , and  $P_{rr}$  are above the signal power spectra. As explained in Appendix C.2, this is due to coupling between galaxy shot noise and clustering at large scales, where the variance is larger than at small scales and therefore these shot noise spectra are significantly boosted compared to the  $\bar{n}^{-1}$  contribution.

The middle panel of Fig. 5.9 shows the expected constraints on  $f_{NL}$  when only  $\delta_g$  is used, or when reconstructed modes are also incorporated. The right panel shows the ratio of  $\sigma(f_{NL})$  in these two cases. The improvement in  $\sigma(f_{NL})$  is negligible at the lowest  $K_{min}$  we consider, which corresponds to  $\delta_g$  being measured on all scales resolvable within the survey volume ( $K_{min} = K_f$ ). However, a larger improvement

is seen when  $K_{\min}$  is assumed to be higher: for  $K_{\min} = 0.02h \text{ Mpc}^{-1}$ , for example,  $\sigma(f_{\text{NL}})$  improves by around 15% when reconstructed modes are used.



*Fig. 5.10* The analog of the right panel of Fig. 5.9, with a variety of (mostly artificial) modifications to the forecasts. There is no improvement in  $\sigma(f_{\text{NL}})$  when  $\delta_r$  is neglected at  $K < K_{\min}$ , indicating that the inclusion of  $\delta_r$  at  $K < K_{\min}$  drives the improvement. Greater improvements are achieved for higher galaxy number density or if  $k_{\max}$  can be increased by a factor of 2, with milder changes if the fiducial  $b_2$  value is set to zero or shot noise on  $P_{\text{rr}}$  and  $P_{\text{gr}}$  is neglected.

To determine the origin of this behavior, we show several modifications of this forecast in Fig. 5.10. In particular, when reconstructed modes with  $K < K_{\min}$  are not included, there is no improvement of  $\sigma(f_{\text{NL}})$ , indicating that these modes are entirely responsible for the improvement. Therefore, DESI is not powerful enough to allow for cosmic variance cancellation between  $\delta_g$  and  $\delta_r$  at the same scales; rather, the primary use of reconstruction is to access scales ( $K < K_{\min}$ ) where  $\delta_g$  cannot be directly measured. This naturally explains why the improvement of  $\sigma(f_{\text{NL}})$  grows for higher  $K_{\min}$ . While the absolute values of  $\sigma(f_{\text{NL}})$  are not impressive at such high  $K_{\min}$  – at  $K_{\min} = 0.02h \text{ Mpc}^{-1}$ , for example,  $\sigma(f_{\text{NL}}) \approx 50$  without reconstruction and 40 with reconstruction – the improvement comes “for free”, without requiring any other datasets.

The other curves in Fig. 5.10 illuminate other aspects of this forecast. Increasing  $\bar{n}$  to an unrealistically high value of  $10^2 \text{ Mpc}^{-3}$  improves the  $\delta_g$ -only forecast by roughly 10% (not shown), and also increases the improvement on  $\sigma(f_{\text{NL}})$  from including  $\delta_r$ , indicating that shot noise is a limiting factor in this improvement. Simply neglecting  $P_{\text{rr,shot}}$  and  $P_{\text{gr,shot}}$  has a similar effect, clarifying that shot noise in the galaxy power spectrum itself is comparatively less important than in these other spectra.

Also, we see the same type of change if we alter the fiducial value of  $b_2$ . As mentioned at the end of Sec. 5.2.3, if  $P_{\text{gg}} \gg P_{\text{gg,shot}}$  (as it is here), then  $P_{\text{rr}}/N_{\text{GG}} \propto (1 + b_2/b_1)^2$ , so increasing  $b_2$  from  $-0.3$  to  $0$  boosts the signal to noise on the reconstructed modes.

This would lead to a larger improvement if not for the large contribution of  $P_{\text{rr,shot}}$ . Boosting  $k_{\text{max}}$  by a factor of 2 leads to a better  $\sigma(f_{\text{NL}})$  improvement at low  $K_{\text{min}}$ . This change lowers the Gaussian reconstruction noise  $N_{\text{GG}}$ , but also raises  $P_{\text{rr,shot}}$  and  $P_{\text{gr,shot}}$  by different amounts, and the combination of these changes ends up slightly boosting the constraining power of  $\delta_{\text{r}}$ .

It may seem counterintuitive that  $\delta_{\text{r}}$  adds anything at all to our forecasts, since the reconstruction noise and shot noise on  $P_{\text{rr}}$  are much larger than  $P_{\text{rr}}$  itself: one would expect such large noise to lead to a low cross-correlation coefficient between  $\delta_{\text{r}}$  and  $\delta_{\text{g}}$ , and also make it difficult to extract information from the auto spectrum of  $\delta_{\text{r}}$ . However, the presence of a cross shot noise contribution to  $P_{\text{gr}}$  alters this picture, contributing to the  $\delta_{\text{r}}\text{-}\delta_{\text{g}}$  cross-correlation coefficient and altering the structure of the covariance matrix. While it is not trivial to see in the Fisher matrix expression in Eq. (5.53), the net effect is to enhance the information content of  $\delta_{\text{r}}$  with respect to  $f_{\text{NL}}$ . Reference [165] reached a similar conclusion when examining cosmic variance cancellation between different line intensity maps, noticing that lowering the cross shot noise contribution led to worsened constraints on  $f_{\text{NL}}$ .

### MegaMapper

Our results for the MegaMapper-like survey are shown in Fig. 5.11. For the low- $z$  bin, the signal to reconstruction noise on the reconstructed modes is higher than for DESI, thanks to a combination of higher  $\bar{n}$ , higher  $k_{\text{max}}$ , and higher bias, and the signal-to-shot-noise ratio is also correspondingly smaller. This leads to a greater improvement in  $\sigma(f_{\text{NL}})$  when reconstructed modes are included. The left panel of Fig. 5.12 shows that, like DESI, this improvement comes not from cosmic variance cancellation, but from reconstructed modes with  $K < K_{\text{min}}$ , where we assume that  $\delta_{\text{g}}$  cannot be directly measured. We see large changes if  $\bar{n}$  is boosted or  $P_{\text{gr,shot}}$  and  $P_{\text{rr,shot}}$  are neglected, indicating that shot noise is a limiting factor in this bin. Changing the fiducial  $b_2$  from 1.1 to 0 reduces the usefulness of the reconstructed modes for the same reason that changing  $b_2$  increased their usefulness for DESI.

We see rather different behavior in the high- $z$  bin. There, we find that the reconstruction noise is of the same order as  $P_{\text{rr}}$  while the shot noise contribution is much greater than the signal, and the shot noise contribution to  $P_{\text{gr}}$  is also greater than the signal. Despite this, the improvement in  $\sigma(f_{\text{NL}})$  is larger than for the low- $z$  bin, reaching 50% at  $K_{\text{min}} = K_{\text{f}}$ . The right panel of Fig. 5.12 shows that the improvement is the same whether or not we include modes of  $\delta_{\text{r}}$  with  $K < K_{\text{min}}$ , and therefore, cosmic variance cancellation between  $\delta_{\text{g}}$  and  $\delta_{\text{r}}$  is solely responsible for the change in  $\sigma(f_{\text{NL}})$ .

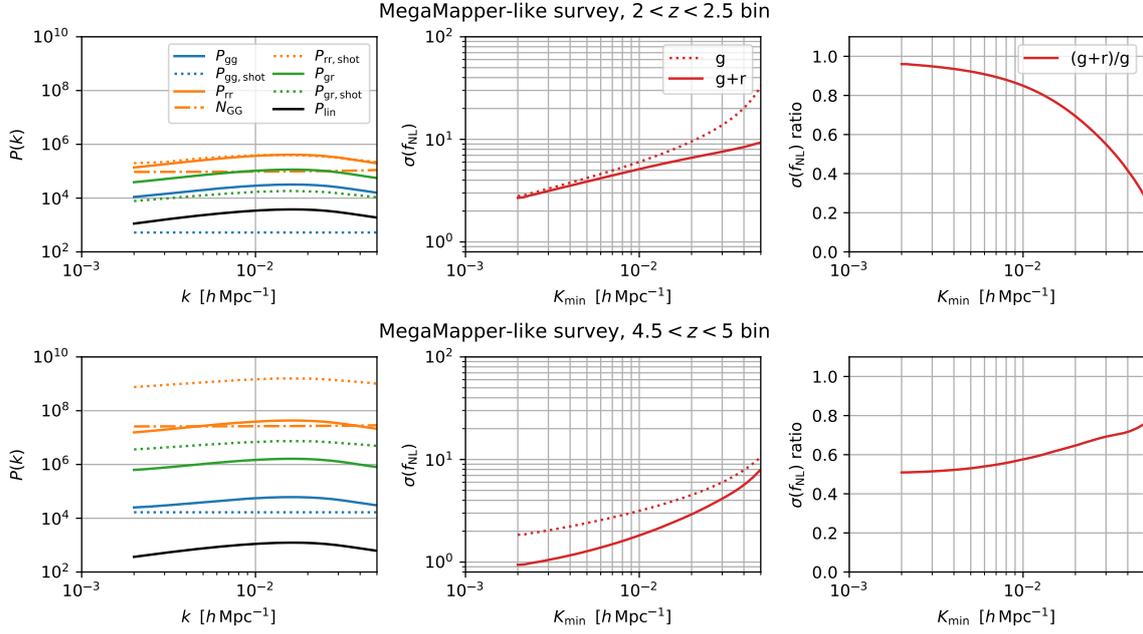


Fig. 5.11 As Fig. 5.9, for low-redshift (*top panels*) and high-redshift (*bottom panels*) bins of a MegaMapper-like survey. The former has greater signal to noise on reconstructed modes than DESI, leading to a greater improvement in  $\sigma(f_{\text{NL}})$  when these modes are included in the forecast. For the latter, the shot noise contributions to  $P_{\text{rr}}$  and  $P_{\text{gr}}$  are comparatively much larger, leading to a different scale-dependence for the improvement in  $\sigma(f_{\text{NL}})$ .

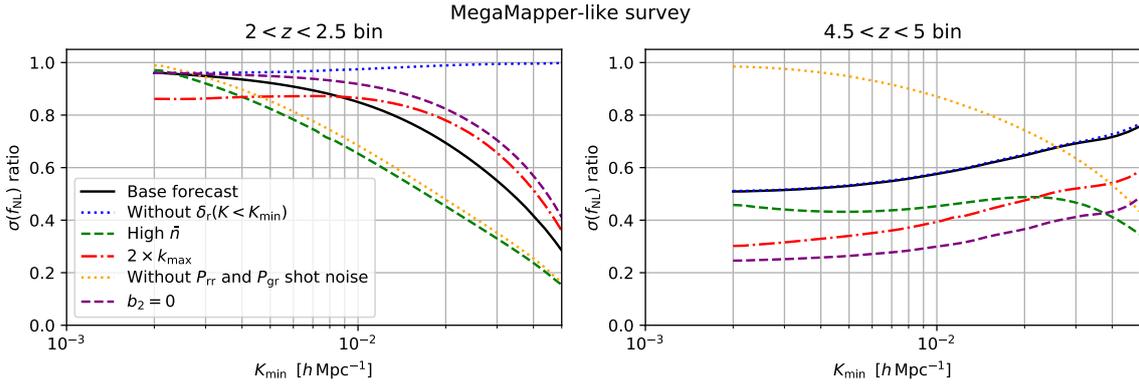


Fig. 5.12 Modifications to the base MegaMapper forecasts. For the low-redshift bin (*left panel*), as for DESI, the improvement in  $\sigma(f_{\text{NL}})$  is driven mostly by modes of  $\delta_{\text{r}}$  with  $K < K_{\text{min}}$ , with further improvement possible for higher galaxy number density. For the high-redshift bin (*right panel*), neglecting  $\delta_{\text{r}}$  at  $K < K_{\text{min}}$  makes no difference, indicating that for lower  $K_{\text{min}}$  values, cosmic variance cancellation between  $\delta_{\text{g}}$  and  $\delta_{\text{r}}$  at the same  $K$  is driving the improvement in  $\sigma(f_{\text{NL}})$ . There are several ways to obtain greater improvements, as discussed in the main text.

We also see from Fig. 5.12 that the low number density ( $\bar{n} = 2 \times 10^{-5} \text{ Mpc}^{-3}$ ) in the high- $z$  bin is not a huge limiting factor, with only a modest change if we use a much larger number density. This is because the reconstruction noise remains comparable to  $P_{\text{rr}}$  even for a much denser survey, while further improvements are possible for a higher  $k_{\text{max}}$  but the same number density. If  $P_{\text{rr,shot}}$  and  $P_{\text{gr,shot}}$  are ignored, the results revert to the same situation as the low- $z$  bin, with only slight gains in  $\sigma(f_{\text{NL}})$  possible for low  $K_{\text{min}}$ . Finally, if  $b_2$  is changed from 17 to 0, there is significantly more improvement in  $\sigma(f_{\text{NL}})$ : the amplitudes of  $P_{\text{rr}}$  and  $P_{\text{gr}}$  are reduced, but the relative uncertainty on  $f_{\text{NL}}$  from marginalizing over  $b_2$  is also reduced, and the latter effect wins.

## PUMA

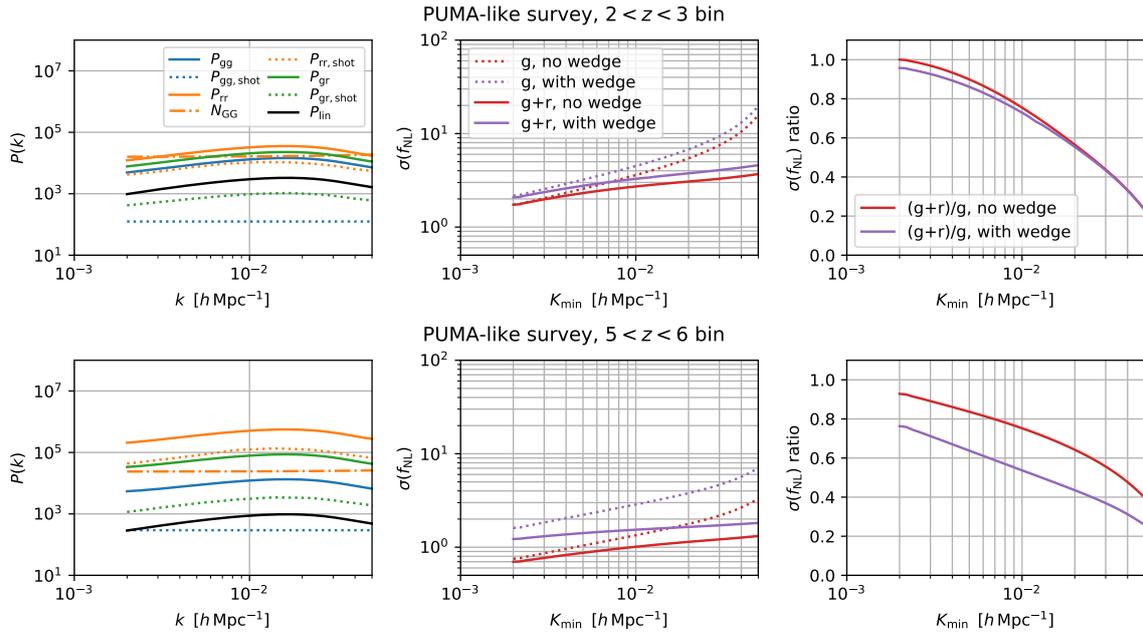
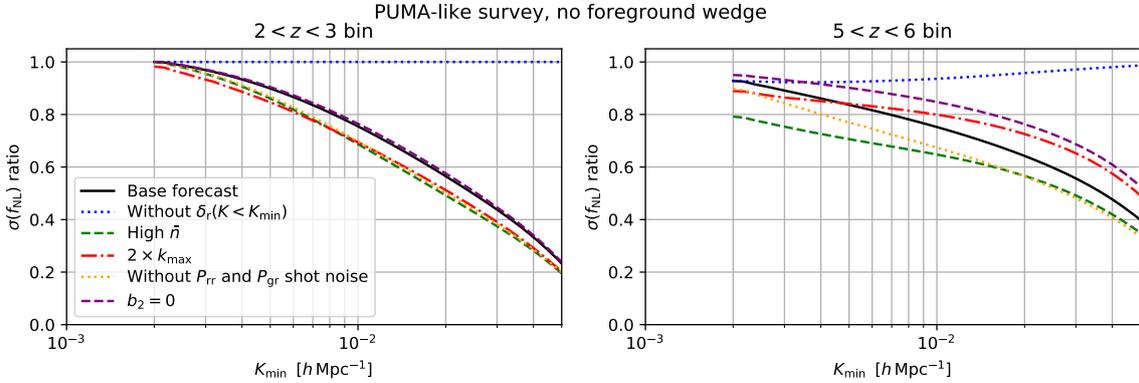


Fig. 5.13 As Fig. 5.9, for low-redshift (*top panels*) and high-redshift (*bottom panels*) bins of a PUMA-like survey, treating the 21 cm brightness temperature in the same way as  $\delta_g$  in our other forecasts, and translating thermal noise on the brightness temperature into an effective tracer number density for computing shot noise. We show  $\sigma(f_{\text{NL}})$  either neglecting or incorporating the effects of the 21 cm foreground wedge; at high  $z$ , the benefit to  $\sigma(f_{\text{NL}})$  from including reconstructed modes is greater in the presence of the wedge, since there are fewer  $\delta_g$  modes that can be directly measured in that case. The results for the low-redshift bin are similar to those for MegaMapper, while larger improvements in  $\sigma(f_{\text{NL}})$  are possible at higher redshift.

We show results for the PUMA-like survey in Fig. 5.13, either neglecting or including the effects of the foreground wedge. Note that the left panels in Fig. 5.13 only show

noise curves corresponding to the no-wedge case. As for the other surveys, we assume an isotropic  $K_{\min}$  for  $\delta_g$  in Fig. 5.13; we show results for a cutoff on  $K_{\parallel}$ , which are qualitatively similar to those in Fig. 5.13, in Appendix C.5.2.

For both redshift bins, the shot noise in  $C^{\text{gg}}$ ,  $C^{\text{rr}}$ , and  $C^{\text{gr}}$  is below the signal. However, the reconstruction noise is high enough in the low-redshift bin that the effect of reconstructed modes on  $\sigma(f_{\text{NL}})$  is similar to DESI and the low- $z$  MegaMapper bin, with the vast majority of the extra constraining power coming from reconstructed modes with  $K < K_{\min}$  (see the left panel of Fig. 5.14). The impacts of taking a higher  $k_{\text{max}}$  or tracer number density (equivalent to thermal noise in the interferometer) would only be mild.



*Fig. 5.14* Modifications to the base PUMA forecasts, neglecting the foreground wedge. In the low- $z$  bin, modes of  $\delta_r$  with  $K < K_{\min}$  are entirely responsible for the improvement in  $\sigma(f_{\text{NL}})$ , with most other modifications having little effect. In the high- $z$  bin, the blue dotted curve demonstrates that the  $\sigma(f_{\text{NL}})$  improvement comes from a combination of low- $K$  modes of  $\delta_r$  and cosmic variance cancellation at higher  $K$ . The improvement would get better if the thermal noise could be reduced (which maps onto a higher  $\bar{n}$  in these forecasts).

Meanwhile, in the high- $z$  bin, the improvement in  $\sigma(f_{\text{NL}})$  arises from a combination of low- $K$  reconstructed modes and cosmic variance cancellation between  $\delta_g$  and  $\delta_r$ . There is greater improvement in the presence of the wedge, as reconstruction helps to recover modes that would otherwise be lost. This improvement is around 20% at the lowest  $K_{\min}$ , and increases as more  $\delta_g$  modes are lost, implying that reconstruction will be extremely useful for single-tracer constraints on  $f_{\text{NL}}$  from PUMA or other high- $z$  intensity mapping. The right panel of Fig. 5.14 shows that lower thermal noise would lead to further improvements, while a lower value of  $b_2$  would worsen the results due to a lowering of the signal to noise on  $P_{\text{rr}}$ .

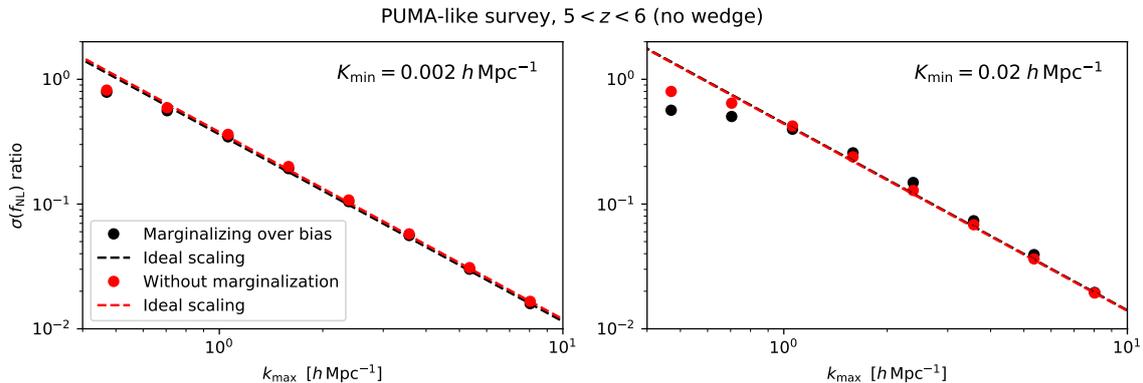
### ...and beyond

To demonstrate how the constraints on  $f_{\text{NL}}$  scale for surveys with extremely low shot noise and reconstruction noise, we also examine forecasts for the PUMA high-redshift bin where  $k_{\text{max}}$  is artificially increased, assuming that our quadratic bias model is valid to arbitrarily high  $k$ .<sup>13</sup> We take the galaxy number density to infinity in these forecasts, to prevent shot noise from becoming the limiting factor. In this case, we expect the uncertainty on  $f_{\text{NL}}$  to scale like the inverse of the signal to noise on the reconstructed modes (see Sec. 5.4.2). In turn, in this limit, the signal to noise scales like  $k_{\text{max}}^{3/2}$  because the reconstruction noise spectrum  $N_{\text{GG}}$  becomes proportional to the number of modes with  $k_{\text{min}} < k < k_{\text{max}}$  (see Eq. C.63).

In Fig. 5.15, we show the ratio of  $\sigma(f_{\text{NL}})$  from g+r or g-only forecasts as a function of  $k_{\text{max}}$  for two representative values of  $K_{\text{min}}$ . We indeed find that as the signal to noise on reconstructed modes is increased, the improvement on  $\sigma(f_{\text{NL}})$  also increases, with the unmarginalized forecasts quickly satisfying the expected scaling. (Marginalization over bias parameters causes small deviations from this scaling.) This demonstrates the huge increases in constraining power that are possible in principle for a survey with high galaxy number density and many small-scale modes whose correlations can be used in reconstruction. We have also numerically verified that the  $\sigma(f_{\text{NL}})$  ratio stays flat with increasing  $k_{\text{max}}$  if the noise in  $P_{\text{gg}}$  is taken very high, or if zero cross-correlation between  $\delta_{\text{g}}$  and  $\delta_{\text{r}}$  is assumed, further demonstrating that the scaling seen in Fig. 5.15 arises from the joint constraining power between  $\delta_{\text{g}}$  and  $\delta_{\text{r}}$  measured from the same volume.

---

<sup>13</sup>In practice the quadratic bias model will break down at sufficiently high  $k$ , but a theoretical framework such as the response function formalism (e.g. [29, 28]) may allow the use of higher  $k_{\text{max}}$ , with suitable modifications of the reconstruction procedure. We leave this topic to future work.



*Fig. 5.15* The ratio of  $\sigma(f_{\text{NL}})$  for the  $\delta_{\text{g}} + \delta_{\text{r}}$  and  $\delta_{\text{g}}$ -only forecasts for the high- $z$  PUMA bin, where  $\bar{n}$  is taken to infinity and  $k_{\text{max}}$  is artificially increased, assuming that our quadratic bias model is valid to arbitrarily high  $k$ . The left and right panels correspond to two representative values of  $K_{\text{min}}$ . We show forecasts after marginalizing over bias parameters (*black points*) and without any marginalization (*red points*), along with the expected  $k_{\text{max}}^{-3/2}$  scaling (*dashed lines*, each normalized to the highest- $k_{\text{max}}$  point plotted). We find the unmarginalized curves quickly approach the ideal scaling, while the marginalized forecasts show small deviations from this scaling. This shows that large increases in constraining power are possible in principle for surveys with very high number density and a large allowed value of  $k_{\text{max}}$ .

## 5.5 Discussion

The results presented here can be compared to other methods either utilizing reconstruction and/or combining a tower of  $n$ -point correlation functions. Compared to most methods proposed in the literature, this work presents an optimal quadratic estimator to reconstruct the large-scale mode. As explained in Sec. 5.4.4, in principle this reconstructed mode can be used for several (cosmological) applications, and here we only explored  $f_{\text{NL}}$  as an application of interest. When comparing this work with previous works, the main question is if the amount of information captured in the statistics of the tracer field is fully exploited. While it will be hard to compare methods directly, here we propose some heuristic arguments where we think our methods overlap and where they differ.

As mentioned in the introduction, some publications have aimed to simplify the search for primordial non-Gaussianities by proposing more compressed versions of the full bispectrum [236, 88, 46, 68, 189, 60, 218, 102]. Common to these works is the fact that the information accessed is captured by the 2- and 3-point functions. In this work, besides the 3-point function, the 4-point function is also used and is important in obtaining cosmic variance cancellation. In other words, as shown in Fig. 5.15, significant improvements are possible when some conditions are met that

would not be possible when considering the compressed statistics proposed in these earlier works.

Even if cosmic variance cancellation is not achieved, we generally observe improvements between 20–50%. These numbers are similar to those projected in for example [68] for compressed statistics, but direct comparisons between our method and others are generally difficult. In the method presented in this chapter, the improvement can roughly be attributed to being able to access larger scales through the reconstruction, or, when both the linear and reconstructed mode are combined, cosmic variance cancellation. The projected improvement on the amplitude  $f_{\text{NL}}$  from compressed statistics is the result of adding the bispectrum information on top of the power spectrum. For a detailed comparison we would need to carefully associate every mode with improved signal-to-noise side by side for the two different methods. Although this would be interesting by itself, as it would help us understand to what extent these methods are overlapping or how they complement one another, we will leave this to future work.

The paper which our work has most in common with is [218], which discusses the information content of a joint analysis of the two point function and squeezed three- and four-point functions. This work has several commonalities with our analysis. To perform forecasts, [218] uses the squeezed-limit position-dependent power spectrum as a field, in an approach that is quite similar to our long-wavelength mode reconstruction. The author also makes similar arguments for how sample variance cancellation can significantly influence and improve constraints.

However, there are also many important differences to our approach. Most importantly, the specific squeezed-limit power spectrum picture in [218] is discussed as a tool to enable better forecasting of joint 2, 3 and 4-point analyses of local non-Gaussianity, rather than a practical data analysis method. In contrast, our method has been proposed as an analysis method and estimator to rapidly jointly analyze 2, 3 and 4-point functions, that is not only computationally tractable, but has been tested (to some extent) on simulations.

There are also significant differences in the details of the methodology. Our reconstruction quadratic estimator can infer the long-wavelength mode from mode-pairs that are not much smaller than the mode to be reconstructed; in contrast, [218] always operates in the squeezed limit when analyzing the position-dependent power spectrum. While it is expected that the majority of information about local non-Gaussianity in the 3 and 4-point functions is contained in very squeezed shapes, it is not clear that non-squeezed shapes do not contribute to long-wavelength mode reconstruction and hence sample variance cancellation. On the other hand, we note that in our analysis method we combine all quadratic estimator mode pairs into

one long-wavelength mode estimate; in contrast, [218] shows that additional sample variance cancellation can be obtained when treating each mode pair (or position-dependent power spectrum bin) as a separate tracer. Although this suggests that further improvements to our method might be possible, the results of [218] suggest that this would only give significant improvements for very high  $k$  and very low noise, beyond the capabilities of next-generation surveys.

Finally, shortly before the completion of this work, in a follow-up to [157], [156] presented results relating to reconstruction of large-scale density modes using biased tracers, although without discussing the application to constraining non-Gaussianity. While the core of this work is similar (using a quadratic estimator as proposed by [91]), here we explicitly account for the mode-coupling from higher-order biasing (which is non-negligible) in our estimator and compare theoretical estimates of the reconstruction noise, including bi- and trispectrum shot noise with additional contributions from primordial non-Gaussianity, to simulations. Reference [156] include observations on the light-cone in their formalism, and also include the effect of redshift space distortions up to second order in the linear density, which we neglect in this work (although see Sec. 5.4.3 for a discussion of the impact of the Kaiser term).

## 5.6 Conclusions

In this chapter, we have further developed a method for reconstructing modes of the cosmic density field using a quadratic estimator. This estimator extracts information about (typically) large-scale modes from correlations between smaller-scale modes, similar to standard methods for CMB lensing reconstruction. We have improved upon the estimator introduced in [91] by incorporating nonlinear biasing and local-type primordial non-Gaussianity, up to second order in the linear density field. At this order, there are several distinct sources of couplings between small-scale modes of the tracer density field, with amplitudes (i.e. bias coefficients) that are unknown a priori. We have found that an estimator based on the mode-coupling due to isotropic growth of the perturbations results in the lowest noise on reconstructed modes, and have enumerated the various multiplicative biases that will accompany the output of this estimator. We have also applied this estimator (along with those based on large-scale bulk flows and tidal interactions) to halos in  $N$ -body simulations, verifying that the results agree with analytical predictions.

In the course of this study, we have identified that it is crucial to include the shot noise contribution to the covariance between directly-observed tracer modes and reconstructed modes when performing an analysis. The shot noise not only

adds a white noise contribution to the tracer power spectrum itself (in the case that the tracers Poisson-sample the density field, which we assume here), but also adds noise to the reconstructed-mode power spectrum and the cross-spectrum with the tracer modes. For sufficiently low tracer number density, this contribution can actually overwhelm the reconstruction noise from the quadratic estimator, and the cross spectrum alters the correlation coefficient between the tracer and reconstructed modes. We self-consistently include these features in our forecasts.

We have carried out forecasts that apply this formalism to several upcoming large-scale structure surveys: the emission-line galaxy survey from DESI [7], the high- $z$  dropout survey envisioned in the MegaMapper proposal [90, 232], and the 21 cm line-intensity survey from the PUMA proposal [19, 261], treated like a galaxy survey with effective number density derived from PUMA’s thermal noise model. Examining the expected errorbars on the power spectrum of the reconstructed modes for  $K < 0.02h \text{ Mpc}^{-1}$ , we find that these errorbars are several times larger than the signal for DESI and a high-redshift bin of MegaMapper. The latter is limited by the low number density of tracers, leading to a high shot noise contribution to the reconstruction noise, while the former’s high reconstruction noise is sourced both by shot noise and a low number of modes used in the reconstruction (i.e. low  $k_{\text{max}}$ ). In the other forecasts, we find that high-S/N reconstructions of the large-scale density power spectrum can be obtained, with the caveat that this spectrum comes with multiplicative biases with known shapes but unknown amplitudes.

We have also computed the expected improvement in constraints on the amplitude of local-type primordial non-Gaussianity,  $f_{\text{NL}}$ , arising from analyzing reconstructed modes along with directly-observed tracer modes. For DESI and low- $z$  bins of MegaMapper and PUMA, the improvement arises solely from being able to access reconstructed modes with  $K < K_{\text{min}}$ , where  $K_{\text{min}}$  denotes the minimum wavenumber at which we assume tracer modes can be measured (for systematics obscuring tracer modes with  $K < K_{\text{min}}$  but not affecting the modes used for reconstruction). On the other hand, for a high- $z$  bin of MegaMapper, the improvement in  $f_{\text{NL}}$  constraints arises solely from cosmic variance cancellation between tracer and reconstructed modes at the same wavenumbers, similar to what can happen with different tracer populations or tracer-lensing cross-correlations [246, 179, 237, 165]. For a high- $z$  bin of PUMA, the  $\sigma(f_{\text{NL}})$  improvement comes from a combination of cosmic variance cancellation and reconstructed modes alone. Generally, cosmic variance cancellation depends on having a sufficiently high cross-correlation coefficient between the tracer and reconstructed modes, but this depends on shot noise in a somewhat complicated way, due to the aforementioned cross shot noise contribution.

The improvement in  $\sigma(f_{\text{NL}})$  also depends on the assumed value of  $K_{\text{min}}$ , so we have plotted the expected constraints as a function of  $K_{\text{min}}$ . In general, reconstructed modes improve  $\sigma(f_{\text{NL}})$  by tens of percents: for example, at  $K_{\text{min}} = 0.01h \text{ Mpc}^{-1}$ ,  $\sigma(f_{\text{NL}})$  improves by a few percent for DESI, 15% and 40% for the low- $z$  and high- $z$  MegaMapper bins we consider, and at least 20% for both  $z$ -bins of PUMA, depending on what is assumed for the 21 cm foreground wedge. We have also shown that in the limit of zero shot noise, and if our quadratic bias model were valid to arbitrarily high  $k$ ,  $\sigma(f_{\text{NL}})$  scales like  $k_{\text{max}}^{-3/2}$ , reflective of the number of small-scale modes used for reconstruction.

There are several possible ways that this work could be extended. For example, we have neglected redshift-space distortions, but they should clearly be incorporated in advance of applying this technique to data. One could also consider applying reconstruction to photometric surveys, after an assessment of the impact of photometric redshift errors on the results. It would be interesting to see how things change if one were to consider the bias model from [238], based on shifted versions of bias operators designed to more fully incorporate large-scale displacements. Given that shot noise can be a limiting factor, the application of weighting schemes (such as [247]) could potentially further lower the shot noise level and improve the reconstruction performance, although these schemes can be difficult to implement in practice; along similar lines, one could apply a different weighting to better extract the scale dependent signal given by  $f_{\text{NL}}$  (see, e.g., [218]). Finally, one could consider investigating nonlinear response functions [29, 28] as a way to increase the number of small-scale modes that could be used in the quadratic estimator. Overall, we expect there to be many applications for reconstructed modes beyond constraints on local-type non-Gaussianity, and we therefore advocate for this reconstruction procedure as a useful tool to increase the scientific returns of upcoming large-scale structure surveys.

## Conclusions and Outlook

### 6.1 Conclusions

For many years we have been exploiting only power spectra of cosmological observables to constrain our cosmological concordance model. Recently, we have started exploring the universe by probing non-Gaussian statistics. The main focus of this thesis was to study methods for a more robust use of such non-Gaussian statistics.

In the first part of the thesis, Chapters 3 and 4, we focused on CMB lensing analyses. Indeed, in recent years, combined analyses of CMB lensing and galaxy clustering have proven powerful for constraining cosmology. They will gain even more importance in the future, due to recent theoretical and computational development for enhanced modelling and analysis, as well as huge datasets from upcoming CMB and large-scale structure surveys. A primary goal is to test our cosmological concordance model. To this end, we will have to be careful about systematic effects.

In Chapter 3 we presented the latest ACT lensing maps, using data from seasons 2014–2015, cross-correlated with CMASS-BOSS galaxies. In addition, we developed a new method for creating foreground-free CMB lensing maps with minimal impact on signal-to-noise, the symmetric estimator. This is important especially for ground-based experiments like ACT, as standard methods usually significantly increase the noise in the CMB lensing estimator. We then demonstrated this technique by applying it to combined multifrequency tSZ-cleaned data from ACT+Planck to create foreground-free CMB lensing maps.

Several other methods exist for dealing with foregrounds in CMB lensing. In Chapter 4 we explored how we can combine these to obtain more robust estimators for CMB lensing analyses. First, we created new estimators by composing geometric and multifrequency data. In particular, we applied the bias hardening technique to the symmetric estimator. Then, we linearly combined these with other CMB lensing

estimators to obtain the best combination of estimators that minimizes the bias, with only a modest noise penalty. We finally applied our methods to simulations for an SO-like experiment, finding that we can reach regimes where the bias on the CMB lensing amplitude is negligible with respect to the noise. This will be useful for percent and sub-percent level real CMB lensing analyses with current and upcoming data.

The first two thirds of the thesis dealt only with two dimensional non-Gaussian information, as encoded in the observed CMB. The last part of this thesis tackled the topic of studying the non-Gaussian statistics of galaxy tracers, a notoriously computationally intensive task. Non-Gaussian statistics in the large-scale structure, for example, could come from non-linear gravitational evolution, and from local primordial non-Gaussianity. Understanding the latter is of key importance, as a detection of  $f_{\text{NL}} \geq 1$  would rule out single-field slow-roll inflation. Local primordial non-Gaussianity shows up in the observed galaxy field through a scale-dependent bias on large-scales, and a modulation of the power at small scales.

By using quadratic methods, similar to the ones of CMB lensing, we can reconstruct an estimate of the matter field on large scales from a galaxy field, in the same volume. By combining the reconstructed field with the galaxy field for a power spectrum analysis, we can indirectly use three and four-point function information. Indeed, the reconstructed field is a two-point function in the input galaxy field; therefore, the reconstructed-galaxy, and reconstructed-reconstructed power spectra are three and four-point functions in the galaxy field, respectively. This method relies on relatively fast algorithms for the reconstruction of the linear field and calculation of power spectra, as opposed to dealing directly with bispectra and trispectra (though this is true only for the squeezed limits). From the combination of the galaxy and reconstructed fields, we showed forecasts on the error bars for local non-Gaussianity. In some cases we are able to tighten constraints significantly, with respect to some galaxy power-spectrum-only analysis. This shows that the method is promising for future searches for primordial non-Gaussianity.

## 6.2 Outlook

Applying the foreground combined cleaning methods of Chapters 3, 4 to current and future data analyses will be interesting. In particular, rerunning the optimization for the ACT survey specifications will be important for understanding how much the best combination depends on the particular settings. Foreground combined cleaning methods will be relevant also for polarization-dominated surveys such as CMB-S4: even if in general foregrounds are less of a problem, in cross-correlation temperature

is still important for smaller scales, and so the methods developed in this thesis will be relevant. Exploring the use of alternative simulations or theoretical modelling of foregrounds (e.g. through the halo model), will be an important topic to explore too, to test the robustness of the results. Finally, an interesting question is whether we can develop a more general foreground-free estimator that does not rely on the combination or composition of known estimators.

The large-scale matter density field reconstruction method seems promising. But, to apply it to data (e.g. from the SPHEREx satellite), we will have to understand how redshift-space distortions and, for some surveys, photometric redshift errors will affect reconstruction. Next steps could include applications to  $N$ -body simulations, and the addition of more realism, e.g. stellar contamination that might be present in observed data or masking. Finally, it would be interesting to explore the method for other cosmological parameters aside from  $f_{\text{NL}}$ .

## Appendices of Chapter 3

### A.1 CMB map pre-processing for lensing reconstruction

In this appendix, we describe in more detail the preprocessing of the ACT CMB maps which are used in the lensing reconstruction process.

The ACT raw maps are made available as four map splits  $D_{A,f,j}$ ,  $j \in \{1, 2, 3, 4\}$  with the same signal but independent instrumental noise contributions through the time-interleaved splitting scheme described in [10] and [62], for each frequency  $f$  and instrumental array  $A$ . For the D56 region, data are from seasons 2014 and 2015 and observations of the sky are made from the following combinations of array-frequency  $(A, f)$ : (PA1-2014, 150), (PA2-2014, 150), (PA1-2015, 150), (PA2-2015, 150), (PA3-2015, 150), (PA3-2015,98), where only the dichroic PA3 array includes observations at both 98 GHz and 150 GHz. For the BN region, the data are from season 2015 only, for the combinations  $(A, f)$ : (PA1-2015, 150), (PA2-2015, 150), (PA3-2015, 150) and (PA3-2015,98). Here, (PA3-2015,150), for example, corresponds to a map made using measurements from the 150 GHz channel of the PA3 detector array collected during the 2015 observing season.

The temperature maps that enter the ACT+*Planck* tSZ-free lensing maps are pre-processed and coadded (with appropriate tSZ deprojection) as described in [170]. All other maps (i.e. temperature maps for the ACT-only lensing maps and the polarization maps) are pre-processed and co-added as follows:

1. To reduce noise and bias from radio sources and to make subsequent Fourier transforms well-behaved, we use source-subtracted maps (see [62, 10]). Some residuals are left in these at the locations of bright compact sources; these are in-painted within each split using the catalog and maximum-likelihood method

described in [170], *i.e.*, we fill holes around compact sources with a constrained Gaussian realization. These holes of 6 arcminute radius are inpainted jointly for  $T$ ,  $Q$ ,  $U$ . The algorithm used follows the brute-force approach presented in [44]. We then use these splits to obtain a co-added map  $D_{A,f}$  using maps of the inverse white-noise variance in each pixel as well as two sub-splits  $D_{A,f,1} = \sum_{j=1,2} D_{A,f,j}$  and  $D_{A,f,2} = \sum_{j=3,4} D_{A,f,j}$  with independent noise. We use these two sub-splits to obtain an estimate of the 2D Fourier-space noise power spectrum  $N_{A,f}(\vec{l})$ , by taking the difference between the mean auto-spectrum of each sub-split and the mean cross-spectrum between the sub-splits, and subsequently smoothing it.

2. We apply an apodized mask to each map which restricts our analysis to the well-crosslinked region used for power spectrum measurements in [62, 10]. To account for pixelization effects, we deconvolve the pixel window function from each map in 2D Fourier space.
3. We next combine the various maps  $D_{A,f}$  into a single CMB map  $M$  on which the lensing reconstruction is performed, for each of  $T$ ,  $Q$  and  $U$ . Unlike in previous work where a real-space coaddition was used [256], we now co-add the maps in 2D Fourier space (since this is more optimal for multifrequency data with different beams) as follows:  $M(\vec{l}) = B_{A_c, f_c}(l) \sum_{(A,f)} w_{A,f}(\vec{l}) D_{A,f}(\vec{l}) B_{A,f}^{-1}(\ell)$  where

$$w_{A,f} = \frac{N_{A,f}^{-1}(\vec{l}) B_{A,f}^2(\ell)}{\sum_{(A,f)} N_{A,f}^{-1}(\vec{l}) B_{A,f}^2(\ell)} , \quad (\text{A.1})$$

are normalized inverse-variance weights. We note that here a deconvolution of the harmonic-space beam  $B_{A,f}(\ell)$  is performed for each array, and finally a convolution to a common map beam  $B_{A_c, f_c}(l)$  is reapplied; the choice of this beam does not matter since it is deconvolved later. This weighting scheme ignores correlations of the noise between arrays. Only the dichroic arrays (PA3,150 GHz) and (PA3, 98 GHz) have substantial ( $\approx 40\%$ ) noise correlations on the scales considered in this work. While this choice of weighting is sub-optimal, on scales where the (98–150) GHz correlation is important, our measurements are dominated by the CMB signal in the 98 GHz frequency and thus neglecting these correlations will not substantially increase the lensing reconstruction noise.

This procedure, performed separately for each of intensity  $T$  and the  $Q$  and  $U$  polarization Stokes components, results in coadded CMB maps  $M_X$  with

$X \in \{T, Q, U\}$ . We repeat the same operations above on the sub-splits  $D_{A,f,i}$ ,  $i \in \{1, 2\}$  to obtain the corresponding maps  $M_{X,i}$ ,  $X \in \{T, Q, U\}$  from which we obtain an estimate of the experimental noise 2D power  $N_X$ ,  $X \in \{T, Q, U\}$  in the same way as described previously. These noise estimates of the co-added maps are used for optimal weighting in the lensing reconstruction.

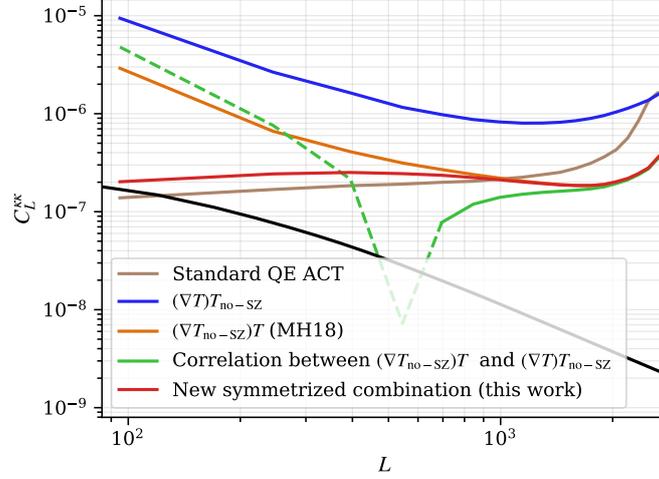
4. While the previously described inpainting procedure removes a large amount of radio source contamination, bright galaxy clusters show up in these maps as decrements due to the thermal Sunyaev-Zel'dovich effect. These add both noise and bias to the lensing estimation, and so we next in-paint a catalog of SZ clusters that have been internally detected. For this catalog, we use confirmed cluster locations inferred from co-add maps that include data up to the 2018 season. From this catalog, we select and inpaint all the clusters with a signal-to-noise ratio greater than 5. The inpainting is performed (only in temperature) within circular holes of 5-arcmin radii using the same method as for the compact sources. A small number of clusters near the edge of the mask that caused problems due to the discontinuous boundary were not inpainted. This is expected to have a negligible impact on our analysis as the number of such clusters is very small, with no particularly bright ones among them. After inpainting, we deconvolve by the common map beam chosen above.

The CMB temperature and polarization maps that result from these steps are used (following filtering and  $E - B$  decomposition) as inputs to our lensing reconstruction pipeline, described in detail in Section 3.

## A.2 Noise properties of the symmetric foreground-cleaned estimator

The goal of this appendix is to illustrate the noise properties of the different lensing estimators used in this work, with particular emphasis on the noise of the new symmetric cleaned estimator that is free of tSZ contamination.

Indeed as explained in the main text, if left untreated, the tSZ induced contamination can bias the results of a low- $z$  galaxy - CMB lensing cross-correlation measurement, by 10%. In combination with cleaned multifrequency data, the lensing estimator we propose below can mitigate these biases, leading to a more robust cross-correlation analysis.



*Fig. A.1* The noise power per mode for the temperature-only estimator for different cases in the D56 region. This plot shows how the symmetric cleaned estimator presented in this work lowers the noise compared to the asymmetric estimator. The green curve shows the cross-noise between the two different asymmetric estimators, with negative values in dashed. The anti-correlation of the noise on large scales between the two different asymmetric estimators leads to a cancellation in the optimal co-add of these that results in the red noise curve for our new symmetric cleaned estimator, which recovers the forecast performance in MH18.

The estimated lensing convergence map in real space from a fixed polarization combination  $XY$  for CMB maps is [e.g., 124]:

$$\hat{\kappa}^{XY}(\hat{n}) = \int \frac{d^2\vec{L}}{(2\pi)^2} e^{i\vec{L}\cdot\hat{n}} \hat{\kappa}^{XY}(\vec{L}), \quad (\text{A.2})$$

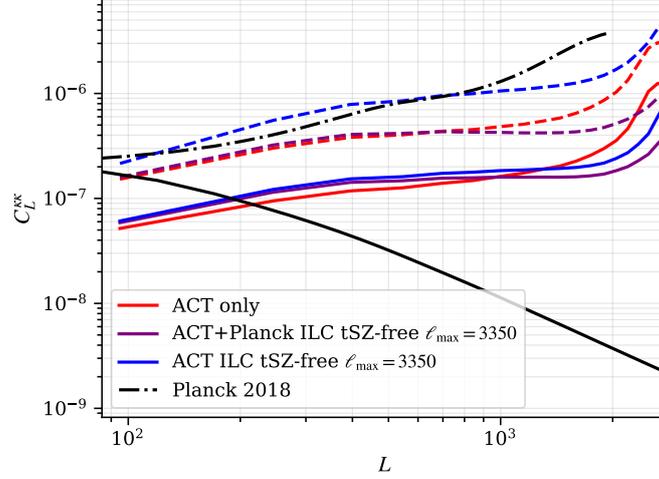
with

$$\hat{\kappa}^{XY}(\vec{L}) = -A_L^{XY} \int d^2\hat{n} e^{-i\hat{n}\cdot\vec{L}} \text{Re}\{\nabla \cdot [\vec{G}^{XY}(\hat{n}) L^{Y*}(\hat{n})]\}, \quad (\text{A.3})$$

where  $XY \in \{T_i T_j, T_i E_j, E_i E_j, E_i B_j\} + i \leftrightarrow j$  with the indices characterizing maps with different data content (e.g. from different experiments or with different component separation techniques),  $A_L^{XY}$  is a normalization to ensure that we recover an unbiased estimate of the convergence field, and  $\vec{G}^{XY}(\hat{n})$  and  $L^{Y*}(\hat{n})$  are filtered versions of CMB maps. The details of these filtered maps can be found in [124].

The normalization is

$$A_L^{XY} = \frac{L^2}{2} \left[ \int \frac{d^2\vec{l}}{(2\pi)^2} (\vec{L} \cdot \vec{l}) W_l^{XY} W_{|\vec{L}-\vec{l}|}^Y f^{XY}(\vec{l}, \vec{L}-\vec{l}) \right]^{-1}, \quad (\text{A.4})$$



*Fig. A.2* The noise power per mode in our maps for different patches for the minimum variance (temperature + polarization) co-add of the CMB lensing maps from this work. Solid colored lines represent D56, dashed lines represent BN and dashed and dotted represent *Planck* 2018. The theory expectation for the signal is shown in black. Our maps are signal dominated for  $L < 100$  in BN and  $L < 200$  in D56.

where  $W_l^{XY}$  is used to obtain  $G^{XY}$  in (A.3),  $W_l^Y$  to obtain  $L^Y$ , with details for these and  $f_{XY}(\vec{l}, \vec{L} - \vec{l})$  that can be found again in [124]. The lensing convergence estimator expands to

$$\hat{\kappa}^{XY}(\vec{L}) = A_L^{XY} \int \frac{d^2\vec{l}}{(2\pi)^2} (\vec{L} \cdot \vec{l}) W_l^{XY} X(\vec{l}) W_{|\vec{L}-\vec{l}|}^Y Y(\vec{L} - \vec{l}). \quad (\text{A.5})$$

The covariance of this estimator,  $N^{XY,WZ}(\vec{L})$  is

$$\begin{aligned} & \langle \hat{\kappa}^{XY}(\vec{L}) \hat{\kappa}^{WZ}(\vec{L}') \rangle_{\text{CMB}} - \langle \hat{\kappa}^{XY}(\vec{L}) \rangle_{\text{CMB}} \langle \hat{\kappa}^{WZ}(\vec{L}') \rangle_{\text{CMB}} = \\ & = (2\pi)^2 \delta_D^{(2)}(\vec{L} - \vec{L}') A_L^{XY} A_L^{WZ*} \int \frac{d^2\vec{l}}{(2\pi)^2} (\vec{L} \cdot \vec{l}) W_l^{XY} W_{|\vec{L}-\vec{l}|}^Y \\ & \quad \times [(\vec{L} \cdot \vec{l}) W_l^{WZ} W_{|\vec{L}-\vec{l}|}^Z C_l^{\bar{X}\bar{W}} C_{|\vec{L}-\vec{l}|}^{\bar{Y}\bar{Z}} \\ & \quad + (\vec{L} \cdot (\vec{L} - \vec{l})) W_{|\vec{L}-\vec{l}|}^{WZ} W_l^Z C_l^{\bar{X}\bar{Z}} C_{|\vec{L}-\vec{l}|}^{\bar{Y}\bar{W}}]. \end{aligned} \quad (\text{A.6})$$

When the maps involved are identical ( $X = Y$ , e.g. for TT and EE estimators where both fields have the same data), the minimum-variance filters have a simple form as shown in [124] and the estimator can be written in a separable manner (i.e., can be written using sums of products of a function of  $\vec{l}_1$  times a function of  $\vec{l}_2$ ) that allows for fast evaluation with FFTs. Moreover, the estimator variance ( $X = Y = W = Z$  above) has a simple relation to the normalization  $N_L \propto A_L L^2$ . This no longer holds when  $X \neq$

$Y$ . In particular, for our case of interest where we mix maps with different component separation techniques,  $X = T_{\text{no-fg}}$  and  $Y = T_{\text{with-fg}}$ , the minimum variance estimator does not have a simple separable form. MH18 used an approximation to the minimum-variance estimator that consisted of the two different maps being independently Wiener filtered. When the weights in the estimator are not minimum-variance, the relation (assumed in the forecast of that paper) that  $N_L \propto A_L L^2$  no longer holds. The true performance is the orange curve in Figure A.1. However, a simple heuristic extension of the MH18 estimator recovers performance close to what was forecast there: the two asymmetric estimators  $\hat{\kappa}(T_{\text{no-fg}}, T_{\text{with-fg}})$ ,  $\hat{\kappa}(T_{\text{with-fg}}, T_{\text{no-fg}})$  combined in a minimum-variance combination  $\hat{\kappa}_{\text{symm,fgfree}}^{TT} = \sum w_\alpha(\vec{L}) \hat{\kappa}_\alpha(\vec{L})$  with weights given by Eq. (3.10), where  $\alpha \in \{(T_{\text{no-fg}} T_{\text{with-fg}}), (T_{\text{with-fg}} T_{\text{no-fg}})\}$  and  $N^{-1}$  the inverse of the  $2 \times 2$  covariance matrix taking into account the cross-correlation between the two estimators.

In Figure A.1 we show the noise curves for this  $TT$  symmetric cleaned estimator, as well as the asymmetric estimators. In Figure A.2 we show lensing minimum-variance noise curves, which include polarization lensing measurements. These are shown for three different cases that differ in how the  $TT$  estimator is calculated: (a) using the tSZ-free symmetric cleaned estimator with both *Planck* and ACT data combined with ILC (our baseline, in purple); (b) using only ACT data with the  $1/N$  co-adding scheme, and no deprojection of foregrounds (red); and (c) using the tSZ-free symmetric cleaned estimator with only ACT data combined with ILC (blue).

## A.3 Data Availability

The maps and masks used for this analysis are available at [https://lambda.gsfc.nasa.gov/product/act/actpol\\_prod\\_table.cfm](https://lambda.gsfc.nasa.gov/product/act/actpol_prod_table.cfm).

## Appendices of Chapter 4

### B.1 CMB lensing biases for a combination of estimators

#### Combination of estimators

Suppose we observe the lensed CMB at several frequencies  $\nu_i, i \in \{1, \dots, N_f\}$ . We can exploit the frequency and scale dependence of foregrounds to build a quadratic estimator  $Q(\vec{L}) = \hat{\kappa}(\vec{L})$  for extracting the CMB lensing signal by minimizing the impact of such foregrounds

$$Q(\vec{L}) = \sum_{\nu\nu'} \int_{\vec{\ell}_1} \int_{\vec{\ell}_2} g_{\nu\nu'}(\vec{\ell}_1, \vec{\ell}_2) T_\nu(\vec{\ell}_1) T_{\nu'}(\vec{\ell}_2) |_{\vec{\ell}_1 + \vec{\ell}_2 = \vec{L}} \quad (\text{B.1})$$

where we sum over pair of frequencies, as we perform the quadratic reconstruction on a pair of CMB temperature maps observed at such frequencies, and we weight with some weighting function  $g$  (that in principle might not be symmetric in  $\vec{\ell}_1, \vec{\ell}_2$  and  $\nu, \nu'$ ). This weight function depends not only on the observables, but also on choices like the maximum mode used in reconstruction  $\ell_{\max}$ .

We can then impose the usual requirement of obtaining an unbiased estimate of the CMB lensing signal  $\langle Q(\vec{L}) \rangle_{\text{CMB}} = \kappa(\vec{L})$  (with the quadratic estimator applied only on CMB maps) and the minimization of a function  $\mathcal{L}$  that depends monotonically on the variance and on the induced foreground biases. What is then the optimal form of  $g$  in this case?

This is a difficult constrained optimization problem that could be simplified as follows. First, we further decompose the sum over frequencies and filters as

$$Q(\vec{L}) = \sum_{i\nu\nu'} \int_{\vec{\ell}_1} \int_{\vec{\ell}_2} g_i(\vec{\ell}_1, \vec{\ell}_2) a_{i\nu\nu'}(\vec{\ell}_1, \vec{\ell}_2) T_\nu(\vec{\ell}_1) T_{\nu'}(\vec{\ell}_2) |_{\vec{\ell}_1 + \vec{\ell}_2 = \vec{L}} \quad (\text{B.2})$$

where now we are summing over weighting functions  $g_i$ , independent from frequencies, and we absorb the frequency dependence in coefficients  $a_{i\nu\nu'}$  that mix the frequency space and the CMB temperature multipoles with the CMB lensing space. Then, we make  $a_{i\nu\nu'}$  separable to obtain

$$Q(\vec{L}) = \sum_{i\nu\nu'} \int_{\vec{\ell}_1} \int_{\vec{\ell}_2} a^i(\vec{L}) g_i(\vec{\ell}_1, \vec{\ell}_2) a_\nu^i(\vec{\ell}_1) a_{\nu'}^i(\vec{\ell}_2) T_\nu(\vec{\ell}_1) T_{\nu'}(\vec{\ell}_2) |_{\vec{\ell}_1 + \vec{\ell}_2 = \vec{L}}. \quad (\text{B.3})$$

In this paper we will focus on the case where we fix the CMB temperature multipole weights  $a_\nu^i(\vec{\ell})$  and focus only on varying the lensing weights. Therefore, our optimization problem is to look for optimal coefficients that give an unbiased lensing estimator  $Q$  that minimizes the impact of foreground biases without strong degradation in the signal over noise with the following form

$$Q(\vec{L}) = \sum_i \int_{\vec{\ell}_1} \int_{\vec{\ell}_2} a^i(\vec{L}) g_i(\vec{\ell}_1, \vec{\ell}_2) T_A(\vec{\ell}_1) T_B(\vec{\ell}_2) |_{\vec{\ell}_1 + \vec{\ell}_2 = \vec{L}} \quad (\text{B.4})$$

where  $T_A, T_B$  are two (possibly different) linear combinations of the individual temperature maps at each frequency. The problem of optimizing the estimator in Eq. (B.1) has therefore been decomposed into two steps:

- Choose best combination of frequency CMB temperature data, with respect to some optimisation request.
- Choose best combination of geometric CMB lensing methods, with respect to a similar optimisation request.

In this paper we will focus on the second problem, while [227] focuses on the first.

## Temperature data

We will refer to a generic quadratic estimator  $i$  that takes two temperature maps  $T_{i1}, T_{i2}$ , with  $\mathcal{Q}_i[T_{i1}, T_{i2}](\vec{L})$ . Note, a priori that the estimator is *not symmetric* with respect to its arguments. We now expand each  $T_{ik}$  ( $k \in \{1, 2\}$ ) as  $T_{ik} = T_{\text{CMB}} + T_{f,ik} + T_{n,ik}$ , where the first term is the lensed CMB (equal for each frequency and method of combination), the second is the foreground  $f$  for the map  $i$  in the leg  $k$ , and the last one

is the noise map for the map  $i$  in the leg  $k$ . From now on, we will ignore this last term in the next calculations, keeping in mind that the experimental noise contribution (and large-scale galactic dust too) will only enter the CMB lensing filters used for reconstruction.

We now consider a set  $\mathcal{S}$  of estimators. They could represent even the same estimator with different input maps (e.g. QE appearing once with maps at one frequency, and then at another frequency). From now on, any sum over the index  $i$  will implicitly assume that  $i \in \mathcal{S}$ .

Our goal is to determine the different foreground-bias-induced contributions to a total estimator given by the combination of other estimators

$$\mathcal{Q}_T(\vec{L}) = \sum_i a_i(\vec{L}) \mathcal{Q}_i[T_{i1}, T_{i2}](\vec{L}) \quad (\text{B.5})$$

Using the bi-linearity of the quadratic estimators in their arguments

$$\mathcal{Q}_T(\vec{L}) = \sum_i a_i(\vec{L}) \left[ \mathcal{Q}_i[T_{\text{CMB}}, T_{\text{CMB}}] + \mathcal{Q}_i[T_{f,i1}, T_{\text{CMB}}] + \mathcal{Q}_i[T_{\text{CMB}}, T_{f,i2}] + \mathcal{Q}_i[T_{f,i1}, T_{f,i2}] \right] \quad (\text{B.6})$$

Let's now calculate the autospectrum of  $\mathcal{Q}_T$ , given by:

$$\langle \mathcal{Q}_T(\vec{L}) \mathcal{Q}_T^*(\vec{L}) \rangle - \langle \mathcal{Q}_T(\vec{L}) \rangle \langle \mathcal{Q}_T^*(\vec{L}) \rangle = \sum_{ij} a_i(\vec{L}) a_j^*(\vec{L}) \langle \mathcal{Q}_i(\vec{L}) \mathcal{Q}_j^*(\vec{L}) \rangle - \langle \mathcal{Q}_T(\vec{L}) \rangle \langle \mathcal{Q}_T^*(\vec{L}) \rangle. \quad (\text{B.7})$$

To obtain this, we have to consider the spectrum obtained from crossing the CMB lensing estimator  $i$  with  $j$ :<sup>1</sup>

$$\begin{aligned} \langle \mathcal{Q}_i[T_{i1}, T_{i2}] \mathcal{Q}_j[T_{j1}, T_{j2}] \rangle &= \langle \mathcal{Q}_i[T_{\text{CMB}}, T_{\text{CMB}}] \mathcal{Q}_j[T_{\text{CMB}}, T_{\text{CMB}}] \rangle \\ &+ \langle \mathcal{Q}_i[T_{\text{CMB}}, T_{\text{CMB}}] (\mathcal{Q}_j[T_{f,j1}, T_{\text{CMB}}] + \mathcal{Q}_j[T_{\text{CMB}}, T_{f,j2}]) \rangle \\ &+ \langle \mathcal{Q}_j[T_{\text{CMB}}, T_{\text{CMB}}] (\mathcal{Q}_i[T_{f,i1}, T_{\text{CMB}}] + \mathcal{Q}_i[T_{\text{CMB}}, T_{f,i2}]) \rangle \\ &+ \langle \mathcal{Q}_j[T_{\text{CMB}}, T_{\text{CMB}}] \mathcal{Q}_i[T_{f,i1}, T_{f,i2}] \rangle + \langle \mathcal{Q}_i[T_{\text{CMB}}, T_{\text{CMB}}] \mathcal{Q}_j[T_{f,j1}, T_{f,j2}] \rangle \\ &+ \langle (\mathcal{Q}_i[T_{f,i1}, T_{\text{CMB}}] + \mathcal{Q}_i[T_{\text{CMB}}, T_{f,i2}]) (\mathcal{Q}_j[T_{f,j1}, T_{\text{CMB}}] + \mathcal{Q}_j[T_{\text{CMB}}, T_{f,j2}]) \rangle \\ &+ \langle \mathcal{Q}_i[T_{f,i1}, T_{f,i2}] \mathcal{Q}_j[T_{f,j1}, T_{f,j2}] \rangle. \end{aligned} \quad (\text{B.8})$$

<sup>1</sup>Note that we have dropped all terms of the form  $\langle \mathcal{Q}_i[T_{f,i1}, T_{\text{CMB}}] \mathcal{Q}_j[T_{f,j1}, T_{f,j2}] \rangle$ , which indeed vanish. To see this, consider Taylor expanding the lensed CMB field in powers of the convergence, schematically we have  $T_{\text{CMB}} \sim T_{\text{CMB}}^0 + \kappa_{\text{CMB}} T_{\text{CMB}}^0 + \dots$ , where  $T_{\text{CMB}}^0$  is the unlensed CMB. Since the unlensed CMB is uncorrelated with the convergence and the foregrounds,  $\langle \mathcal{Q}_i[T_{f,i1}, T_{\text{CMB}}] \mathcal{Q}_j[T_{f,j1}, T_{f,j2}] \rangle \propto \langle T_{\text{CMB}}^0 \rangle = 0$ . Likewise, all terms with three  $T_{\text{CMB}}$ 's will vanish, since the unlensed CMB has a null bispectrum.

Recalling that  $\mathcal{Q}_i[T_{\text{CMB}}, T_{\text{CMB}}] = \kappa_{\text{CMB}} + \text{noise}$ , the above simplifies to:

$$\begin{aligned}
\langle \mathcal{Q}_i[T_{i1}, T_{i2}] \mathcal{Q}_j[T_{j1}, T_{j2}] \rangle &= \langle \kappa_{\text{CMB}} \kappa_{\text{CMB}} \rangle \\
&+ \langle \kappa_{\text{CMB}} (\mathcal{Q}_i[T_{f,i1}, T_{f,i2}] + \mathcal{Q}_j[T_{f,j1}, T_{f,j2}]) \rangle \\
&+ \langle (\mathcal{Q}_i[T_{f,i1}, T_{\text{CMB}}] + \mathcal{Q}_i[T_{\text{CMB}}, T_{f,i2}]) (\mathcal{Q}_j[T_{f,j1}, T_{\text{CMB}}] + \mathcal{Q}_j[T_{\text{CMB}}, T_{f,j2}]) \rangle \\
&+ \langle \mathcal{Q}_i[T_{f,i1}, T_{f,i2}] \mathcal{Q}_j[T_{f,j1}, T_{f,j2}] \rangle,
\end{aligned} \tag{B.9}$$

where we have neglected the Gaussian contribution from the noise, or the  $N^0$  component. On the first row we have the CMB lensing signal, and on the others the foreground biases. These can be deconstructed in the following way, useful for calculation:

- the second row can be viewed as a cross correlation between the CMB lensing convergence field and the map coming from applying the estimators to the foreground maps. Following [202, 278, 229], we call this **Primary Bispectrum** term  $P_{ij}$ .
- the third row constitutes a secondary contraction (i.e. an integral) over the primary term, also known as **Secondary Bispectrum**. Following [229] we will calculate the secondary bias as a cross correlation, by expanding  $T_{\text{CMB}} = T_{\text{CMB},0} + T_{\text{CMB},1} + \mathcal{O}(\kappa^2)$ , and calculate it as

$$\begin{aligned}
S_{ij} &= \langle (\mathcal{Q}_i[T_{f,i1}, T_{\text{CMB},0}] + \mathcal{Q}_i[T_{\text{CMB},0}, T_{f,i2}]) (\mathcal{Q}_j[T_{f,j1}, T_{\text{CMB},1}] + \mathcal{Q}_j[T_{\text{CMB},1}, T_{f,j2}]) \rangle \\
&+ \langle (\mathcal{Q}_i[T_{f,i1}, T_{\text{CMB},1}] + \mathcal{Q}_i[T_{\text{CMB},1}, T_{f,i2}]) (\mathcal{Q}_j[T_{f,j1}, T_{\text{CMB},0}] + \mathcal{Q}_j[T_{\text{CMB},0}, T_{f,j2}]) \rangle
\end{aligned} \tag{B.10}$$

- the last row comes from the autospectrum of the foregrounds only. This includes a signal part plus a Gaussian one. We subtract the latter by creating randomizing the phases in the foreground map,<sup>2</sup> with the same power spectrum as the non Gaussian ones. The difference is the **Trispectrum** term  $T_{ij}$ .

In the end, for a combination of estimators the total foreground induced CMB lensing bias is

$$B_{TT} = \vec{a} \cdot (T + P + S) \cdot \vec{a} \tag{B.11}$$

## Including polarization data

A practical lensing analysis will ideally extract all the signal-to-noise available from observations, and this will include the use of polarization data to extract the lensing

<sup>2</sup>Note, this is not equivalent to "Gaussianize" a field. Homogeneity and gaussianity of a field imply that real and imaginary parts of the field are independent, and the phases are randomly drawn from a uniform distribution on  $[0, 2\pi]$ . The converse is in general not true [149].

signal [130]

$$\hat{\kappa} \equiv \mathcal{Q}_{\text{MV}} = \sum_{XY \in \mathcal{E}} \alpha_{XY} \mathcal{Q}_{XY}[X, Y] \quad (\text{B.12})$$

such that  $\sum_{XY \in \mathcal{E}} \alpha_{XY} = 1$  to obtain an unbiased estimator, and  $XY \in \mathcal{E} = \{TT, TE, TB, EE, EB\}$ . The combined power spectrum using temperature and polarization data is

$$\begin{aligned} C &= \vec{\alpha}^T \mathbf{C} \vec{\alpha} = \sum_{XY, WZ} \alpha_{XY} \alpha_{WZ} C_L^{XY, WZ} \\ &= \sum_{XY, WZ} \alpha_{XY} \alpha_{WZ} C_L^{\kappa\kappa} + \sum_{XY, WZ} \alpha_{XY} \alpha_{WZ} N_L^{XY, WZ} + \dots \\ &= C_L^{\kappa\kappa} + \sum_{XY, WZ} \alpha_{XY} \alpha_{WZ} N_L^{XY, WZ} + \dots \end{aligned} \quad (\text{B.13})$$

If we want to minimize the Gaussian disconnected noise  $N_L = \sum_{XY, WZ} \alpha_{XY} \alpha_{WZ} N_L^{XY, WZ}$  subject to the constraint of having an unbiased estimator, then we have that

$$\vec{\alpha} = \frac{N^{-1} \vec{e}}{\vec{e}^T N^{-1} \vec{e}} \quad (\text{B.14})$$

where  $\vec{e}$  is a vector of only ones, and  $N^{-1}$  is the inverse, per mode, of the spectrum matrix  $N^{XY, WZ}$  for  $XY, WZ \in \mathcal{E}$ . Correlations among estimators are negligible (e.g. for QE at most at the 10% percent level, see [130]), so we neglect them when calculating the weights, and set the matrix  $N$  to be diagonal. As in the main text, we have that the variance on the estimated amplitude per mode

$$\sigma^2(\hat{A}(\vec{L})) = \frac{\sigma_{\text{MV}}^2}{(C_L^{\kappa\kappa})^2}, \quad (\text{B.15})$$

and the bias on the estimated amplitude per mode becomes

$$b(\hat{A}(\vec{L})) = \sum_{XY, WZ \in \mathcal{S}} \frac{\alpha_{XY} \alpha_{WZ} B_L^{XY, WZ}}{C_L^{\kappa\kappa}} = \frac{\vec{\alpha}^T \mathbf{B} \vec{\alpha}}{C_L^{\kappa\kappa}} \quad (\text{B.16})$$

Our goal is to again calculate the total function

$$\mathcal{L} = \sigma^2(\hat{A}) + f_b^2 b(\hat{A})^2 \quad (\text{B.17})$$

The calculations are very similar to the temperature only case. The results are

$$\begin{aligned} \mathcal{L} = & \int_{\vec{L}} w^2(\vec{L}) \times \sum_{XY,WZ,AB,CD} \left( \frac{\Theta^{XY,WZ,AB,CD}(\vec{L})}{(C_L^{\kappa\kappa})^2} \alpha_{XY}(\vec{L}) \alpha_{WZ}(\vec{L}) \alpha_{AB}(\vec{L}) \alpha_{CD}(\vec{L}) + \right. \\ & \left. + f_b^2 w(\vec{L}) \alpha_{XY}(\vec{L}) \alpha_{WZ}(\vec{L}) \frac{B^{XY,WZ}(\vec{L})}{C_L^{\kappa\kappa}} \times \int_{\vec{L}'} w(\vec{L}') \alpha_{AB}(\vec{L}') \alpha_{CD}(\vec{L}') \frac{B^{AB,CD}(\vec{L}')}{C_{L'}^{\kappa\kappa}} \right) \end{aligned} \quad (\text{B.18})$$

Any  $TT$  part will be given by the combination of temperature estimators (e.g.  $B_{\text{comb},TT} = \vec{a} \cdot B_{TT} \vec{a}$ ). The polarization contributions will be from a single estimator, and we will use the standard quadratic estimator on any polarization pair  $XY \in \{TE, EE, EB, TB\}$ .<sup>3 4</sup>

We now calculate the total bias per mode, as given from equation B.13

$$B = \vec{a} \cdot \mathbf{B} \cdot \vec{a} = \alpha_{TT}^2 B_{TT} + 2\alpha_{TT} \sum_{XY \in \text{pol}} \alpha_{XY} B^{TT,XY} + \dots, \quad (\text{B.19})$$

as  $B^{XY,WZ} = B^{WZ,XY}$ , and we therefore need  $B^{XY,WZ}(\vec{L})$ . To this end, we assume each map to be composed of a CMB contribution and an astrophysical foreground contribution,  $X = \text{CMB} + \text{fg}$  (ignoring the experimental noise component, or any large-scale galactic foreground contribution). We can then write the contributions to the combined temperature and polarization power spectrum as (ignoring correlations of the type  $\langle X_{\text{CMB}} f f \rangle$ ).

$$\begin{aligned} C^{XY,WZ} & \rightarrow \langle \mathcal{Q}_I[X_{I1}, Y_{I2}] \mathcal{Q}_J[W_{J1}, Z_{J2}] \rangle \\ & = \langle \mathcal{Q}_I[X_{\text{CMB}}, Y_{\text{CMB}}] \mathcal{Q}_J[W_{\text{CMB}}, Z_{\text{CMB}}] \rangle \\ & + \langle \mathcal{Q}_I[X_{\text{CMB}}, Y_{\text{CMB}}] (\mathcal{Q}_J[W_{f,J1}, Z_{\text{CMB}}] + \mathcal{Q}_J[W_{\text{CMB}}, Z_{f,J2}]) \rangle \\ & + \langle \mathcal{Q}_J[W_{\text{CMB}}, Z_{\text{CMB}}] (\mathcal{Q}_I[X_{f,I1}, Y_{\text{CMB}}] + \mathcal{Q}_I[X_{\text{CMB}}, Y_{f,I2}]) \rangle \\ & + \langle \mathcal{Q}_J[W_{\text{CMB}}, Z_{\text{CMB}}] \mathcal{Q}_i[X_{f,I1}, Y_{f,I2}] \rangle \\ & + \langle \mathcal{Q}_I[X_{\text{CMB}}, Y_{\text{CMB}}] \mathcal{Q}_J[W_{f,J1}, Z_{f,J2}] \rangle \\ & + \langle (\mathcal{Q}_I[X_{f,I1}, Y_{\text{CMB}}] + \mathcal{Q}_I[X_{\text{CMB}}, Y_{f,I2}]) (\mathcal{Q}_J[W_{f,J1}, Z_{\text{CMB}}] + \mathcal{Q}_J[W_{\text{CMB}}, Z_{f,J2}]) \rangle \\ & + \langle \mathcal{Q}_I[X_{f,I1}, Y_{f,I2}] \mathcal{Q}_J[W_{f,J1}, Z_{f,J2}] \rangle \end{aligned} \quad (\text{B.20})$$

<sup>3</sup>It should be possible to generalize the other non QE estimators, for example bias hardening, to the polarization. For simplicity we will just suppose to have QE to retain maximum signal over noise, as we assume no foregrounds for CMB polarization maps.

<sup>4</sup>We ignore  $ET$ .

where  $I, J$  are shorthands to indicate an estimator for the combinations  $XY, WZ$  respectively. Furthermore, we assume that the polarization data is immune from foregrounds, under the assumption that a point source mask is enough to remove effects from polarized point sources. We want to focus on the part of the CMB lensing estimator that includes polarization data and gives foregrounds bias contributions, under the previous assumption. Let's set  $X \in \{T\}$ ,  $Y \in \{T, E, B\}$ ,  $W \in \{T, E, B\}$ ,  $Z \in \{E, B\}$

$$\begin{aligned}
C^{TY,WZ} &= C^{WZ,TY} \rightarrow \langle \mathcal{Q}_I[T_{I1}, Y_{I2}] \mathcal{Q}_J[W_{J1}, Z_{J2}] \rangle \\
&= \langle \mathcal{Q}_I[T_{\text{CMB}}, Y_{\text{CMB}}] \mathcal{Q}_J[W_{\text{CMB}}, Z_{\text{CMB}}] \rangle + \\
&\quad + \langle \mathcal{Q}_I[T_{\text{CMB}}, Y_{\text{CMB}}] (\mathcal{Q}_J[W_{f,J1}, Z_{\text{CMB}}]) \rangle \\
&\quad + \langle \mathcal{Q}_J[W_{\text{CMB}}, Z_{\text{CMB}}] (\mathcal{Q}_I[T_{f,I1}, Y_{\text{CMB}}] + \mathcal{Q}_I[T_{\text{CMB}}, Y_{f,I2}]) \rangle \\
&\quad + \langle \mathcal{Q}_J[W_{\text{CMB}}, Z_{\text{CMB}}] \mathcal{Q}_I[T_{f,I1}, Y_{f,I2}] \rangle \\
&\quad + \langle (\mathcal{Q}_I[T_{f,I1}, Y_{\text{CMB}}] + \mathcal{Q}_I[T_{\text{CMB}}, Y_{f,I2}]) \times \mathcal{Q}_J[W_{f,J1}, Z_{\text{CMB}}] \rangle
\end{aligned} \tag{B.21}$$

These correlations give the following

$$\langle \kappa_{\text{CMB}} \kappa_{\text{CMB}} \rangle + \langle \kappa_{\text{CMB}} \mathcal{Q}_I[T_{f,I1}, Y_{f,I2}] \rangle + \langle (\mathcal{Q}_I[T_{f,I1}, Y_{\text{CMB}}] + \mathcal{Q}_I[T_{\text{CMB}}, Y_{f,I2}]) \times \mathcal{Q}_J[W_{f,J1}, Z_{\text{CMB}}] \rangle \tag{B.22}$$

Now, the only way to have a primary bispectrum term is if  $Y \in \{T\}$ , given our assumptions, and for our case this is simply

$$\frac{1}{2} P_{TT,I} = \frac{1}{2} \sum_i a_i \mathcal{Q}_i[T_{f,i1}, T_{f,i2}] = \frac{1}{2} \sum_i a_i P_{ii} \tag{B.23}$$

where  $P_{ii}$  is the primary that we have already calculated from  $TT$  data for the temperature  $i$  estimator.

For the secondary bispectrum part,  $W \in \{T\}$ , as we assume no foregrounds in polarization maps. If  $Y \in \{T\}$ , then we have  $S^{TT,TE}$ ,  $S^{TT,TB}$ , and this is (we consider only  $TE$ )<sup>5</sup>

$$S^{TT,TE} = \langle (\mathcal{Q}_I[T_{f,I1}, T_{\text{CMB}}] + \mathcal{Q}_I[T_{\text{CMB}}, T_{f,I2}]) \times \mathcal{Q}_J[T_{f,J1}, E_{\text{CMB}}] \rangle \tag{B.24}$$

---

<sup>5</sup>If  $Y \in \{E, B\}$ ,  $S^{TE,TE}$ ,  $S^{TB,TB}$ ,  $S^{TE,TB}$ ,  $S^{TB,TE}$  we still have contributions, that we here ignore, as the polarization weight for  $TB$  is practically negligible, and for  $TE$  becomes not too important when squared.

and we use again the trick to calculate the secondary by decomposing  $X_{\text{CMB}} = X_{\text{CMB}}^0 + X_{\text{CMB}}^1$ ,  $X \in \{T, E\}$  in powers of  $\kappa$ . To recap, we obtain a symmetric bias matrix  $B^{XY,WZ}$  whose components are<sup>6</sup>

$$B^{TT,TT} = \vec{a} \cdot B_{TT,TT\text{-data}} \cdot \vec{a} \quad (\text{B.25})$$

$$B^{TT,EB} = B^{TT,EE} = \frac{1}{2} \sum_i a_i P_{ii} \quad (\text{B.26})$$

$$B^{TT,TE} = \frac{1}{2} \sum_i a_i P_{ii} + S^{TT,TE} \quad (\text{B.27})$$

$$B^{TT,TB} = \frac{1}{2} \sum_i a_i P_{ii} \quad (\text{B.28})$$

For the optimization, we then take its absolute value and smooth it with a Gaussian as with the  $TT$  case, and calculate

$$B_{TT+\text{pol}} = \vec{\alpha} \cdot B \cdot \vec{\alpha} \quad (\text{B.29})$$

where  $B$  has components  $B^{XY,WZ}$ . In Figure B.1 we show how an example of how minimum variance weights when combining temperature and polarization data change when including bias contributions.<sup>7</sup> Finally, given our assumptions, for the cross-correlation we do not expect to have a significant foreground bias arising from the polarization estimator contributions.

## B.2 Composition of CMB lensing estimators

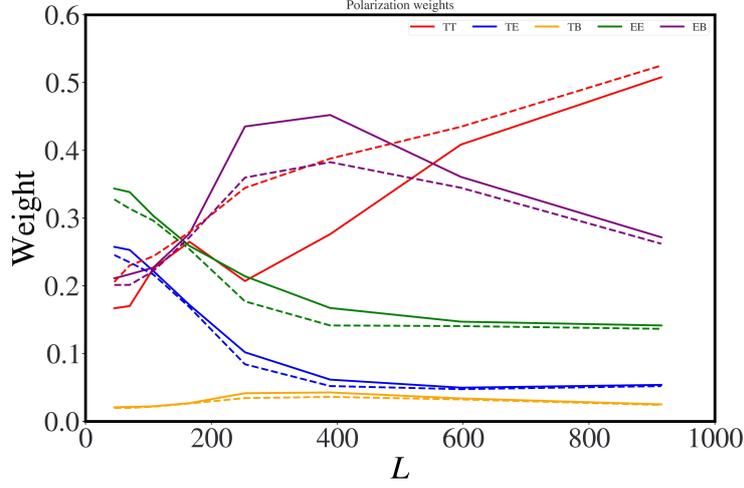
In this section we discuss how we compose, i.e. simultaneously apply, the symmetric estimator and the bias hardening operation. First, we will review the symmetric estimator. Then, we review the bias hardening operation.

### B.2.1 The symmetric estimator

The symmetric estimator is built on a foundation of the gradient cleaning method [169]. The idea behind the gradient cleaned estimator is the following. The standard quadratic estimator for CMB lensing estimation can be obtained from the divergence

<sup>6</sup>We ignore the secondary contraction for  $TETE$  as it enters with a small weight of  $\alpha_{TE}^2$ . We do not consider a secondary for  $TB$ . Even setting  $B^{TT,TB} = 0$  should be fine, given the negligible weight for  $TB$  in the  $MV$  combination.

<sup>7</sup>To calculate the Gaussian lensing noise from polarization data, we use a minimum variance noise from the SO polarization noise frequencies.



*Fig. B.1* Polarization weights change when accounting for the presence of biases. The solid line represents the minimum variance weights for each polarization estimator when including the presence of biases in the optimization, with  $f_b = 4$ . The dashed line is the minimum variance weight when we do not consider the presence of biases. We can see that the combined  $TT$  estimator becomes less important when including biases, and the  $EB$  one is favored.

of  $T_{\text{filt},1} \vec{\nabla} T_{\text{filt},2}$ .  $T_{\text{filt},1}$  is an inverse variance weighted CMB map. On the other hand,  $T_{\text{filt},2}$  is a Wiener filtered CMB map useful for the estimation of the gradient of the unlensed CMB map. Because of diffusion damping, the information coming from the latter saturates after CMB scales at  $l \sim 2000$  [124]. The intuition behind the gradient cleaned estimator is then to use a noisier foreground-free cleaned CMB map  $T_{\text{depr},2}$  in the gradient leg,  $T_{\text{filt},1} \vec{\nabla} T_{\text{depr},2}$ . This results in cleaned cross-correlations with matter tracers, or auto-correlations without trispectrum or primary bispectrum terms [169].

Now, let's write the temperature maps as the superposition of two sets of modes  $T = T_{\text{low}} + T_{\text{high}}$ , where low indicates modes below some scale  $l_c$ , and high otherwise. The problem of the gradient cleaned estimator is that it misses the high – high CMB lensed multipole correlations. This results in a high noise on large scales in the reconstructed CMB lensing map. The symmetric estimator in [72] recovers these modes, by creating a linear combination of gradient cleaned estimators, by taking the divergence of  $aT_{\text{filt},1} \vec{\nabla} T_{\text{depr},2} + bT_{\text{depr},2} \vec{\nabla} T_{\text{filt},1}$ . This can also be written in Fourier space as  $\hat{\kappa}_{\text{symm}} = \int_{\vec{l}} g_{\text{symm}} T_1(\vec{l}) T_{\text{depr},2}(\vec{L} - \vec{l})$ , with formulae that can be found in [72]. The result is a much lower noise-cost in the symmetrised CMB lensing map gaussian noise, compared to gradient cleaning.

### B.2.2 Bias Hardening

Bias hardening [196] methods have already been explored in detail; see for example [226]. Here, we briefly summarize the idea.

If the total temperature map is  $T = T_{\text{CMB}} + s$ , where  $s$  is a foreground source, we have that

$$\langle T(\vec{l})T(\vec{L} - \vec{l}) \rangle = f^\kappa(\vec{l}, \vec{L} - \vec{l})\kappa(\vec{L}) + f^s(\vec{l}, \vec{L} - \vec{l})f(\vec{L}) \quad (\text{B.30})$$

for a statistical average with the CMB lensing convergence field and foreground fixed at some mode  $\vec{L}$ .

Therefore, the estimator for the CMB lensing convergence field, say with  $T_A, T_B$  temperature maps

$$\hat{\kappa}(\vec{L}) = \int_{\vec{\ell}} T_A(\vec{\ell})T_B(\vec{L} - \vec{\ell})g_{AB}(\vec{\ell}, \vec{L} - \vec{\ell}), \quad (\text{B.31})$$

picks up some foreground contribution [196].

The idea behind bias hardening is to write an unbiased estimator for the foreground source

$$\hat{s}(\vec{L}) = \int_{\vec{\ell}} T_A(\vec{\ell})T_A(\vec{L} - \vec{\ell})g_s(\vec{\ell}, \vec{L} - \vec{\ell}). \quad (\text{B.32})$$

One can then write (we omit the argument  $\vec{L}$ )

$$\begin{pmatrix} \langle \hat{\kappa} \rangle \\ \langle \hat{s} \rangle \end{pmatrix} = \begin{pmatrix} 1 & R^{s\kappa} \\ R^{\kappa s} & 1 \end{pmatrix} \begin{pmatrix} \kappa \\ s \end{pmatrix}, \quad (\text{B.33})$$

where  $R^{ab}$  is the response of the estimator  $a$  on the field  $b$ ,  $R^{ab} = \int_{\vec{l}} g_a(\vec{l}, \vec{L} - \vec{l})f_b(\vec{l}, \vec{L} - \vec{l})$ . Inverting this system it then possible to write a weighting function for the bias hardened estimator<sup>8</sup>

$$g_{BH,AB} = \frac{1}{1 - R^{s\kappa}R^{\kappa s}}(g_{AB} - R^{s\kappa}g_s). \quad (\text{B.34})$$

For composing bias hardening with the symmetric estimator, we just then set  $g_{AB} \equiv g_{\text{symm}}$ , applying it on  $T_A, T_B = T, T_{\text{depr}}$ , a map with the foreground, and another without it respectively.<sup>9</sup>

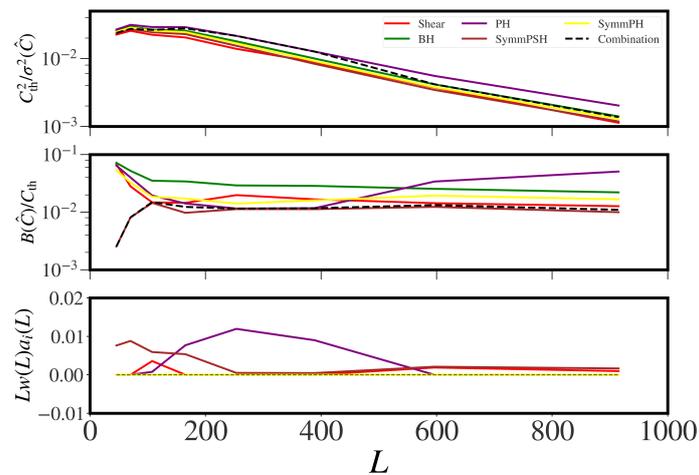
<sup>8</sup>Another way to write the estimator, to easier reconnect to the CMB lensing literature, is the following. Set  $g_{AB} = A_\kappa F_\kappa$ , with some normalisation  $A_\kappa$ , and  $g_s = A_s F_s$ , with some normalisation  $A_s$ . Then,  $F_{BH} = F_\kappa - F_s A_s / A_{\kappa s}$ , where  $A_{\kappa s} = [\int_{\vec{l}} F_\kappa f_s]^{-1}$ . The normalization is  $A_{BH} = [\int_{\vec{l}} F_{BH} f_\kappa]^{-1}$ .

<sup>9</sup>Note that  $g_s$  was built to act on  $T_A, T_A$  only, and not on the deprojected map too. Therefore, when  $g_s$  is applied on  $T_A, T_B$   $T_A \neq T_B$ , it is a slightly suboptimal estimator in terms of variance for the foreground source, although it is still capable of detecting it for removal.

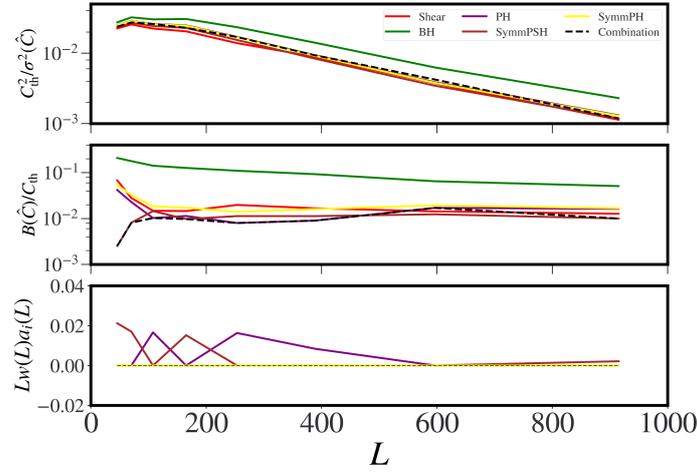
## B.3 Optimization explorations

In this section we present a few details about the optimization results presented in the main text.

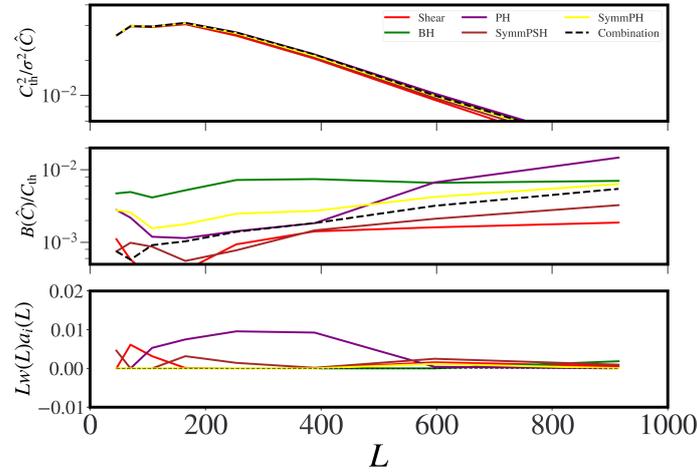
### Per mode results



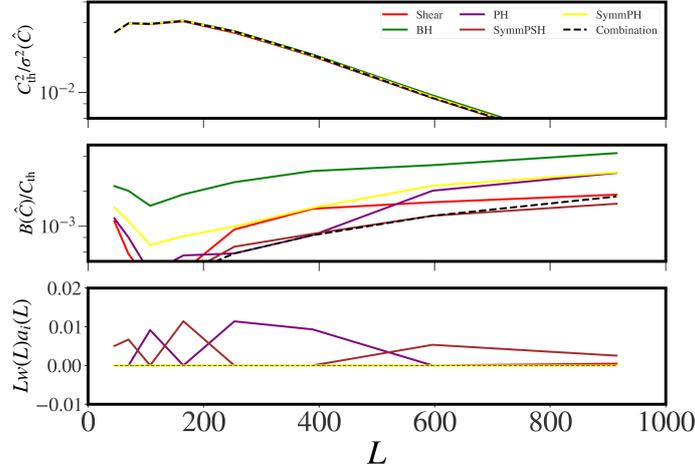
*Fig. B.2* Per mode plot for the case of  $f_b = 1$  for the cross power TT only data optimization. On the top plot we have the signal over noise squared per mode from  $TT$ , then a filtered version, with a Gaussian filter, of the absolute value of the total foreground bias for each estimator. On the bottom, the weights per mode for each estimator.



*Fig. B.3* Per mode plot for the case of  $f_b = 4$  for the cross power  $TT$  only data with an LSST like sample optimization. On the top plot we have the signal over noise squared per mode from  $TT$ , then a filtered version, with a Gaussian filter, of the absolute value of the total foreground bias for each estimator. On the bottom, the weights per mode for each estimator.



*Fig. B.4* As for Figure 4.10, but now showing the case of  $f_b = 1$  for the cross power  $TT$  plus polarization data optimization. It can be seen that, for  $f_b = 1$  where bias and variance are assigned equal importance, the optimization tries to compromise between bias and noise per mode.



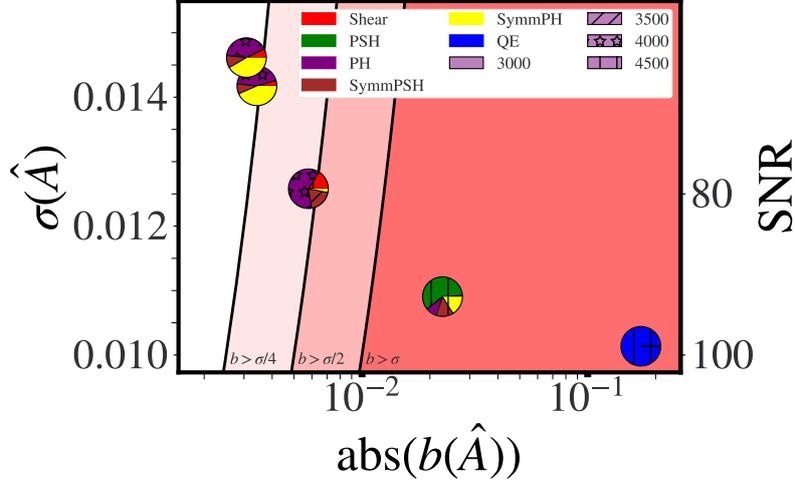
*Fig. B.5* As for Figure 4.10, but now showing the case of  $f_b = 4$  for the cross power  $TT$  plus polarization data optimization. It can be seen that, for  $f_b = 4$  generally, the estimator with lowest bias gets selected in the combination.

### Dependence of biases and optimization on the point source mask

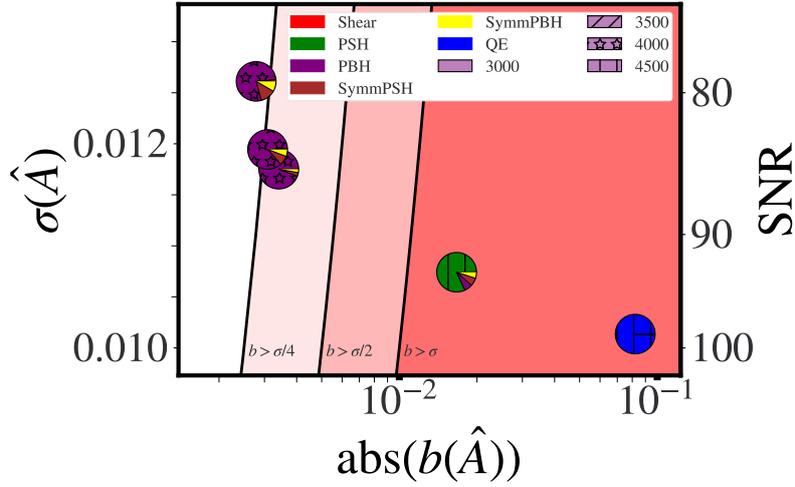
We may run our pipeline with a more aggressive masking approach by running the matched filter for point sources at all the SO frequencies, and not just a single one as in the main text and keeping as our source catalog all the objects found at any frequency. The results of the aggressive masking optimization are in Figure B.6b, to be compared to the standard ones, that for convenience are in Figure B.6a. During the optimization of the more aggressive masking, that covers 8% of the observed sky, we do not consider the change in the available sky fraction, with respect to the standard mask, with 3% masking are. The sky fraction is fixed to the SO one, to  $f_{sky} = 0.4$ . We can see that the noise penalty for high  $f_b$  is lower in the aggressive mask case, compared to the standard one. This can be explained by a change in the biases for each estimator, where now the  $PH$  has smaller biases, but has a good signal over noise compared to other estimators.

## B.4 Lensing validation

We verified our pipeline with two checks. In Figure B.7a we show the lensing validation results, from the mean of the cross-spectrum of 80 Gaussian lensed simulations with the corresponding  $\kappa$  map (note that this is on a small area  $20 \times 20 \text{ deg}^2$  of cut sky so exact agreement with 1 is not expected). While in Figure B.7b we show the analytical Gaussian noise of the CMB lensing power spectrum vs the measured power spectrum of a few estimators when applied to mock Gaussian CMB maps with a total power



(a) Optimization results for the auto correlation for  $TT$  data. The pie charts represent the contribution from each estimator, calculated as  $\int_{\vec{L}} w(L) a_i(L)$ . We can see that for high deprojection, on the extreme left, we need a combination of geometric and multi-frequency deprojection methods.



(b) Optimization results for the auto correlation for  $TT$  data with aggressive mask. The pie charts represent the contribution from each estimator, calculated as  $\int_{\vec{L}} w(L) a_i(L)$ . When using the aggressive mask, we fix the ILC weights to the ones of the baseline mask, and maintain the same fraction of sky. We can see that for high deprojection, the results change compared to the baseline mask. The SymmPH has no trispectrum or primary term, just secondary. While the PH has all of them. The aggressive mask impacts more the trispectrum term, reducing it. The optimization process then prefers to use more PH, compared to the baseline mask, for high deprojection.

*Fig. B.6* On the left, optimization results for our baseline mask used in the main text. On the right, optimization results for a more aggressive mask, with respect to the baseline one. We can see how the results are sensitive to the used mask.

spectrum equal to the total power of the filter used in the estimators (lensed CMB + detector noise + foregrounds).

## B.5 Optimization details

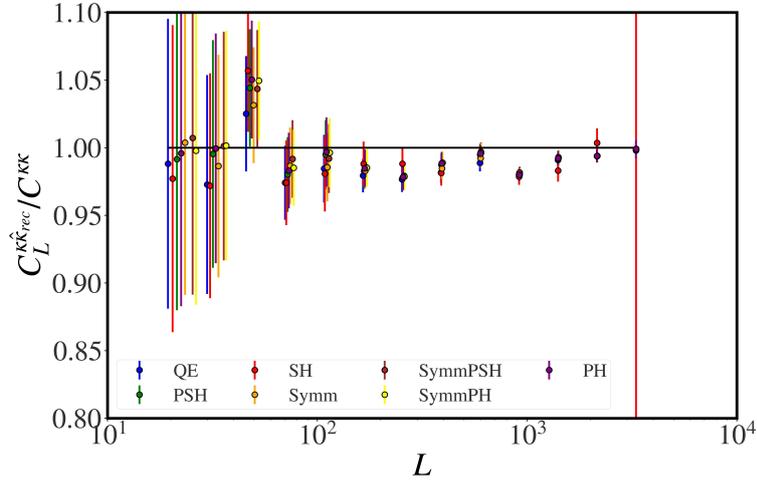
### B.5.1 Choosing among different loss functions

When performing the optimization in practice the variance part will include only the total variance for power spectra, namely cosmic variance and (reconstruction) noise, without the bias contribution. The bias enters separately explicitly in another term, as seen in equation 4.17.

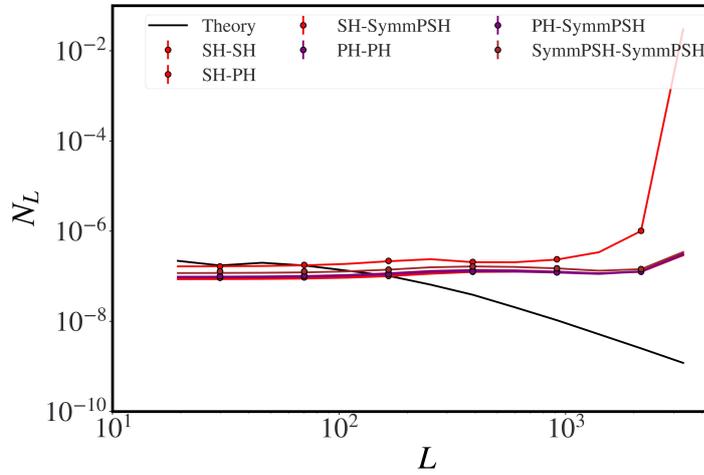
We do not know the real true foreground biases in nature. The biases that we use to calculate the optimal combination in general will arise: from some theory model (e.g. halo model), or from simulations, as in this paper. In both cases, care is required when deciding what to include as a bias in the loss function. Here we list a few possibilities.

- 1 Taking as bias the sum over foregrounds of the absolute values of the different parts of the bias, i.e. the trispectrum, primary and secondary contributions  $\sum_f (|T| + |P| + |S|)$ .
- 2 Taking as the bias the sum over foregrounds  $\sum_f (T + P + S)$ .
- 3 Taking as the bias the absolute value of the sum of biases over foregrounds.  $|\sum_f (T + P + S)|$ .
- 4 Taking as the bias the total bias arising from the sum of the foreground maps  $(T + P + S)_{\text{total}}$ , where total = tSZ + kSZ + CIB + radio.
- 5 Taking as the bias the absolute value of the sum of the foreground maps  $|(T + P + S)_{\text{total}}|$ .
- 6 Taking as the bias  $(|T| + |P| + |S|)_{\text{total}}$ .

Among these choices, 2, 3, 4, and 5 allow for cancellations among the trispectrum, primary and secondary contribution that lead to zero crossing. In particular, 2 and 4 allow for cancellations among estimator biases at the same  $\vec{L}$  or among different  $\vec{L}$ 's in the integral of 4.17. The zero crossing effect and any type of cancellation can lead to very misleading conclusions, as these are highly simulation/model dependent, and



(a) All the estimators have practically unit response. The large red bar on the right comes from the shear estimator, that becomes sub-optimal at small scales.



(b) In solid, the Gaussian reconstruction noise from the theory expression, vs the measured power spectrum in dot of some lensing estimators when applied to a sim map with no lensing (Gaussian random field with power spectrum equal to the total power spectrum). We can see that the dotted points agree with the theory curves.

Fig. B.7 Validation of the pipeline for generating maps, lensing maps, and reconstructing lensing for QE, Shear, SymmPSH, SymmPH, PSH, Symm, evaluated on 80 mock maps,  $20 \times 20 \text{ deg}^2$  each.

depend on the experimental configuration and the  $l_{\max}$ 's of reconstruction. Therefore, using 2 or 4 in equation 4.17 is not a wise choice.<sup>10</sup>

To overcome this, one might want to use one of the other choices, where there are absolute values that do not allow proper zero crossing. In 1 and 6, zero crossings are by construction not allowed, although in 3 and 5 they might be allowed, as there could be cancellations among  $T, P, S$  inside the absolute value. On the other hand, we checked that 1 and 6 will make the optimization process prefer configurations with low  $l_{\max}$ -es (of around 2500-3000) as the bias contribution becomes too dominant, and the statistical significance of the measurement decreases greatly. Therefore, 3 and 5 are the only choices that remains. Between these the more realistic one is 5, as biases are calculated from the map of the sum of the foregrounds: this is what we usually have in a data analysis.

Now, how do we solve for the possible zero crossing problem of 5? One way is to use a regularizer  $g_r(\vec{a}, B^{ij}, \dots)$ : indeed the idea is that we want to not mathematically achieve the best possible solution for a fixed configuration. Ways to regularize are:

1. Smooth the input bias with some kernel, so that a bias at some bin is weighted with neighbouring biases.
2. Ignore the zero crossing. Optimise regardless, and choose by hand configurations that do not have the smallest total function equation 4.17 above, i.e. noise squared plus bias squared.
3. Introduce priors on the weights, such that the cost function increases if weight is given to estimators with zero crossing. It should not change the optimal solution much, but allows for better behaved solutions.

In this work we opt for the simple choice of option 1 for regularization. To summarize, we take  $|(T + P + S)_{\text{total}}|$  and we smooth it with some function  $K$ , such that the input bias is<sup>11</sup>

$$B_{\text{input}} = K(|(T + P + S)_{\text{total}}|) \quad (\text{B.35})$$

<sup>10</sup>Another important point to be made is that with zero crossing sometimes, an estimator that is practically not used, will enter the combination only in one mode, just because its bias in that mode is much smaller than the one of other estimators because of zero crossing. One then gets spikes in the weighting combination, that do not carry big physical meaning, as the zero crossing might depend on the simulation fidelity, on the experimental configuration and analysis choices. This might not give general results. This effect can be mitigated with a regularizer.

<sup>11</sup>We take  $K$  to be a Gaussian,  $K(L) = \frac{1}{Z} \exp\left(-\frac{(L-L_0)^2}{2\sigma^2}\right)$ , with  $Z$  a normalization constant, and we take  $\sigma = 1.5\Delta/\sqrt{8\log 2}$ , with  $\Delta$  the width of the bin edges whose center is  $L_0$ . Then the smoothed version of the bias at  $L_0$  is just  $\sum_{\text{bins}} B(L_{\text{bin}})K(L_{\text{bin}})$ .

In practice we take  $K$  to be a Gaussian with  $\sigma = 1.5$  in band-powers. This choice of the input bias should be a realistic non-optimistic one.<sup>12</sup>

## B.5.2 Optimization algorithm

For the optimization algorithm we use Differential Evolution (DE), a gradient free global optimization algorithm. DE is based on *mutation* and *crossover* steps that look for solutions, and a *selection* step that drives in the right direction of global optimization.<sup>13</sup> We use the implementation of the mystic library [181, 180].<sup>14</sup>

## B.5.3 Estimators over which to optimize

Given  $N_e$  lensing estimators, each calculated for  $N_{l_{max,i}}$  lmaxes, supposing a fixed  $l_{min}$ , we have a total of  $\prod_i N_{l_{max,i}}$  configurations over which to optimize to choose the best ones for a given  $f_b$ . In particular, for a fixed  $N_{l_{max}}$  number of lmaxes for each estimator, the number of configurations is  $N_e \times N_{l_{max}}$ . For a fixed configuration of estimators, we then have  $n_{bins} \times (N_e + 1)$  parameters. As we show in Appendix B.6 this can be reduced to  $n_{bins} \times N_e$ .

For a simple case of seven estimators and eight bins this is  $8 \times 7 = 56$  total parameters over to optimize.

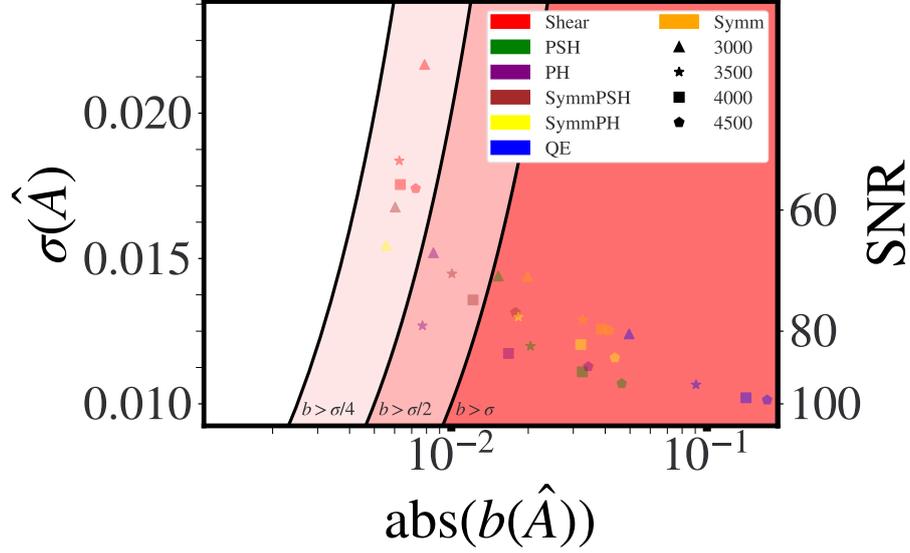
To optimize in a reasonable time (for a fixed configuration a couple of hours), we employ a heuristic rule: we optimize for several  $l_{max}$ 's for each estimator, as show in Figure B.8 for the  $TT$  only case. Then we choose the estimators with configurations nearer the origin in order to conduct further optimization. The idea is that an estimator very far from the origin will have a higher noise or higher bias, or both. Therefore, if for example we intend to minimize the bias, then we will discard the estimators with very high bias and far from the origin, as they will not be useful for the combined optimal estimator. Hence, unless  $f_b = 0$ , we choose to optimize over [Shear, PSH, PH, SymmPSH, SymmPH], with possible lmaxes for each estimator of  $3000 + i500, i = 0, 1, 2, 3$ : this gives 1024 total configurations per  $f_b$ , which is numerically tractable.

---

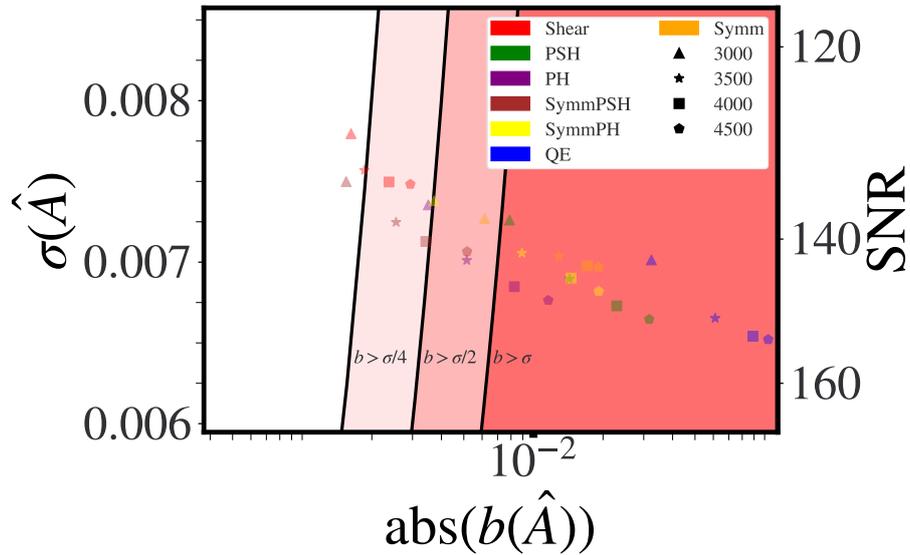
<sup>12</sup>Note, that another way to include uncertainty in foreground biases is by producing a range of simulations, or even better, using some theory foreground model in function of some parameters (e.g. given by halo model) to calculate the variation in the optimization in the weights, and check for the robustness of the optimization in function of varying bias deprojection request.

<sup>13</sup>We check the stability of the optimization procedure by optimizing the same configuration for several  $\mathcal{O}(10)$  times, and we find the final total functional is the same, although some times the optimizer finds different solution with sub-percent differences in the total functional.

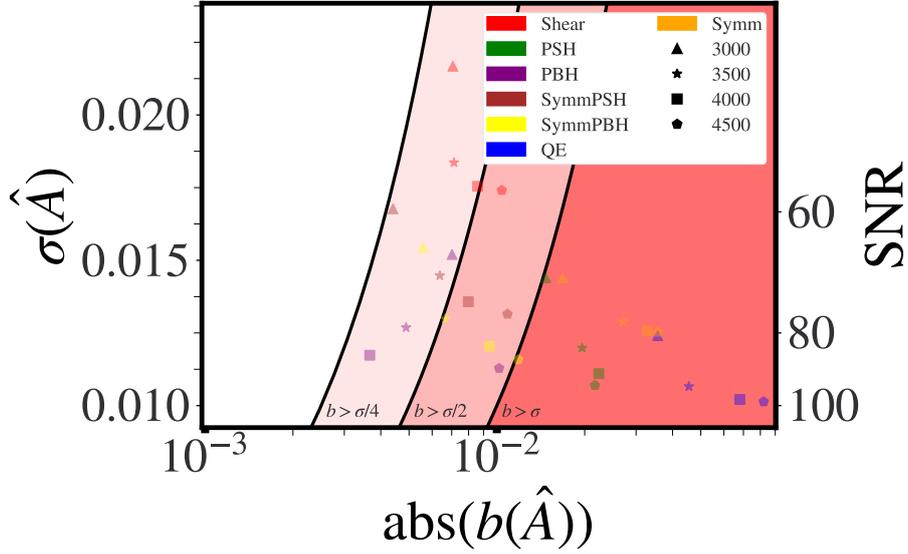
<sup>14</sup><https://github.com/uqfoundation/mystic>.



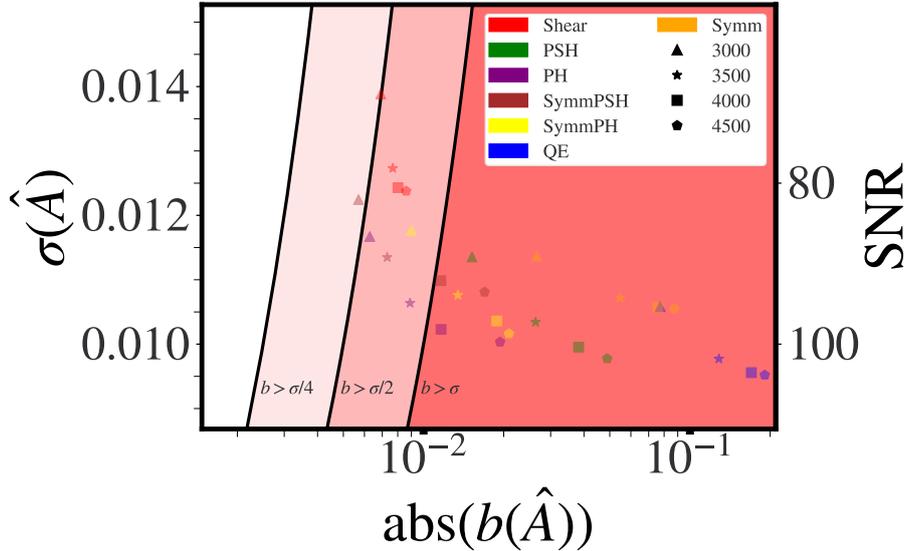
*Fig. B.8* Biases and noises on the CMB lensing amplitude  $A$  for the auto spectrum, with  $TT$  only data at different  $l_{\text{max}}$ es of reconstruction. The bands represent regimes where the bias is greater than a fraction of the noise. We can see in action the bias-noise trade off: as we increase the noise, the biases get reduced, and vice versa.



*Fig. B.9* Biases and noises on the CMB lensing amplitude  $A$  for the auto spectrum, with  $TT$  plus (foreground free) polarization data, for the estimators presented in the main text, at different  $l_{\text{max}}$ es of reconstruction. Thanks to how clean the polarization is, we are easily able to reduce the bias without a high cost in SNR.



*Fig. B.10* Same as Figure B.8 for  $TT$  with a more aggressive masking. In general, the aggressive masking will have more impact on the trispectrum term of the foreground induced CMB lensing bias. This will lead to a change in hierarchy among the biases at different  $l_{\max,TT}$  for the same estimator, e.g. for PH going to  $l_{\max,TT} = 4000$  leads to a lower bias with respect to  $l_{\max,TT} = 3000$ ,



*Fig. B.11* Biases and noises on the CMB lensing amplitude  $A$  for the cross spectrum with an LSST-like sample, with  $TT$  only data, for the estimators presented in the main text, at different  $l_{\max}$  of reconstruction. Here, the bias barely reaches the regime of  $b < \sigma/2$ . Therefore, for this case an optimization is needed to reach a subdominant bias with respect to the noise.

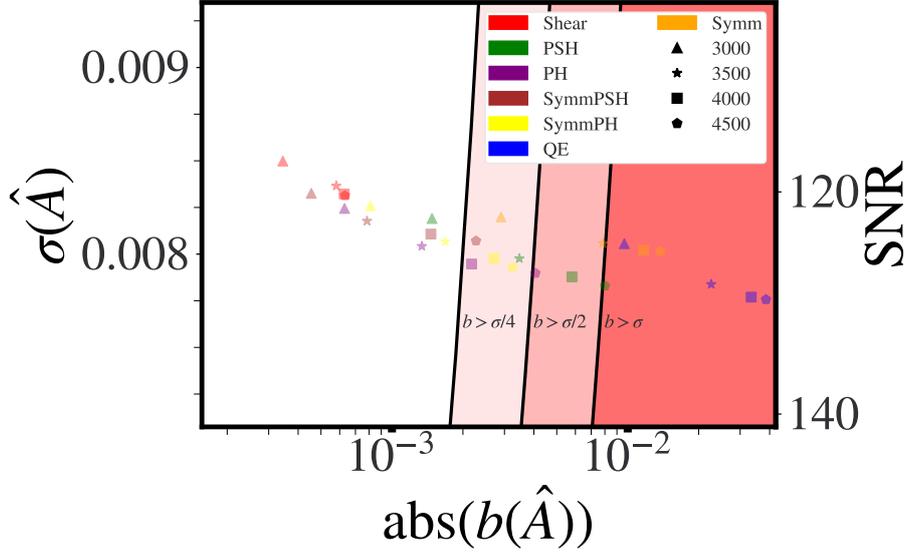


Fig. B.12 Same as Figure B.11 for the cross spectrum with an LSST-like sample, for  $TT$  plus (foreground free) polarization. Thanks to how clean the polarization is, we are easily able to reach a regime where the bias is subdominant compared to the noise for several estimators at several  $l_{\max, TT}$ -es, as the cross-primary term goes as  $\alpha_T T^2$ .

## B.6 Understanding the weights

In this appendix we want to understand the optimal weights analytically for some specific case. Let's write again

$$\begin{aligned} \mathcal{L} = \int_{\vec{L}} w^2(\vec{L}) \frac{1}{4\pi f_{sky}} \sum_{ijmn} \frac{\Theta^{ijmn}(\vec{L})}{(C_L^{\kappa\kappa})^2} a_i(\vec{L}) a_j(\vec{L}) a_m(\vec{L}) a_n(\vec{L}) \\ + f_b^2 w(\vec{L}) a_i(\vec{L}) a_j(\vec{L}) \frac{B^{ij}(\vec{L})}{C_L^{\kappa\kappa}} \int_{\vec{L}'} w(\vec{L}') a_n(\vec{L}') a_m(\vec{L}') \frac{B^{mn}(\vec{L}')}{C_{L'}^{\kappa\kappa}} \quad (\text{B.36}) \end{aligned}$$

Solving this analytically is a very difficult problem.<sup>15</sup> In the case of positive only integrands, e.g. only positive input biases in the optimizer, to minimize  $f$  we need to minimize the integrand at each point, as we do not have cancellations among different  $L$ s. Therefore, if we have the optimal configuration for each bin, then we could imagine to minimize over  $\vec{a}$ , then over  $w$ . Let's then fix the weights combination per mode, and just vary  $w$ . We will use the Lagrange multipliers method. The constraints are

<sup>15</sup>For the case  $f_b = 0$  it is possible to write the minimum variance problem in a matrix form, where the tensor is written as a matrix (e.g. [282]). Or, it is possible to just notice that everything is positive. When  $f_b \neq 0$  the same trick becomes difficult to reapply, and one in general might want to introduce extra constraints.

$\int_{\vec{L}} w(\vec{L}) = 1$ , and  $w(\vec{L}) > 0, \forall \vec{L}$ .

$$\mathcal{L} = \int_{\vec{L}} w^2(\vec{L}) \Theta(\vec{L}) + f_b^2 \left( \int_{\vec{L}} w(\vec{L}) B(\vec{L}) \right)^2 + \lambda \left( 1 - \int_{\vec{L}} w(\vec{L}) \right) + \mu(\vec{L}) (s^2(\vec{L}) - w(\vec{L})) \quad (\text{B.37})$$

where  $\Theta$  is the combined variance per mode divided by the theory squared,  $B$  is the combined bias per mode divided by the theory,  $\lambda$  is a Lagrange multiplier, and  $\mu$  is a Lagrange multiplier function, and  $s$  is a slack variable for the inequality part of the problem. For simplicity, from now on we will call  $w(\vec{L}) = w(\vec{L})$ , and omit the arguments of  $\Theta$  and  $B$ . We note that  $4\pi f_{sky}$  can be absorbed into the definition of  $f_b^2$ , so it can be ignored when obtaining the solution, and substituted again as  $f_b^2 \rightarrow 4\pi f_{sky} f_b^2$ . Basically, a lower  $f_{sky}$  gives larger noise as there are less modes, effectively making the deprojection harder.

When the inequality constraint is not active anywhere

$$w(\vec{L}) = \frac{C_L^2}{\sigma_L^2} \frac{1}{\int_{\vec{L}} \frac{C_L^2}{\sigma_L^2}} + 4\pi f_{sky} f_b^2 \left( \frac{\int_{\vec{L}} \frac{B_L C_L}{\sigma_L^2}}{\int_{\vec{L}} \frac{C_L^2}{\sigma_L^2}} \right) \left( \frac{1 + 4\pi f_{sky} f_b^2 \int_{\vec{L}} \frac{B_L^2}{\sigma_L^2} - 4\pi f_{sky} f_b^2 \frac{\int_{\vec{L}} \frac{B_L C_L}{\sigma_L^2}}{\int_{\vec{L}} \frac{C_L^2}{\sigma_L^2}} \int_{\vec{L}} \frac{B_L C_L}{\sigma_L^2}}{\int_{\vec{L}} \frac{C_L^2}{\sigma_L^2}} \right) \times \left( \frac{C_L^2}{\sigma_L^2} \frac{1}{\int_{\vec{L}} \frac{C_L^2}{\sigma_L^2}} \int_{\vec{L}} \frac{B_L C_L}{\sigma_L^2} - \frac{B_L C_L}{\sigma_L^2} \right) \quad (\text{B.38})$$

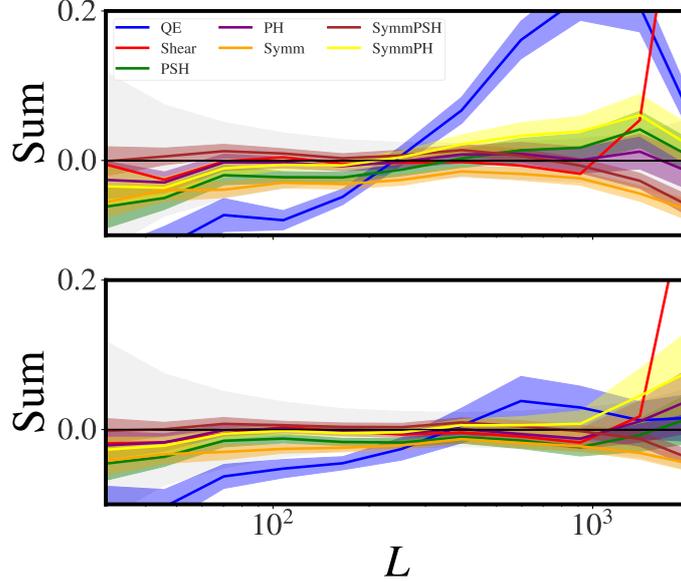
where the quantity in the first big round brackets is the total integrated bias

$$b = \left( \frac{\int_{\vec{L}} \frac{B_L C_L}{\sigma_L^2}}{\int_{\vec{L}} \frac{C_L^2}{\sigma_L^2}} \right) \left( \frac{1 + 4\pi f_{sky} f_b^2 \int_{\vec{L}} \frac{B_L^2}{\sigma_L^2} - 4\pi f_{sky} f_b^2 \frac{\int_{\vec{L}} \frac{B_L C_L}{\sigma_L^2}}{\int_{\vec{L}} \frac{C_L^2}{\sigma_L^2}} \int_{\vec{L}} \frac{B_L C_L}{\sigma_L^2}}{\int_{\vec{L}} \frac{C_L^2}{\sigma_L^2}} \right) \quad (\text{B.39})$$

We can see that this is just the minimum variance solution, with a correction depending on the bias:

$$w(\vec{L}) = w_L^{MV} \left( 1 + 4\pi f_{sky} f_b^2 b \int_{\vec{L}} \frac{B_L C_L}{\sigma_L^2} \right) - 4\pi f_{sky} f_b^2 b \frac{B_L C_L}{\sigma_L^2} \quad (\text{B.40})$$

When the constraint is active somewhere, then we have to substitute the expression for  $w(\vec{L})$  there, and basically  $\mu$  enforces that  $w(\vec{L}) = 0$ , for the specific  $\vec{L}$ . We verify



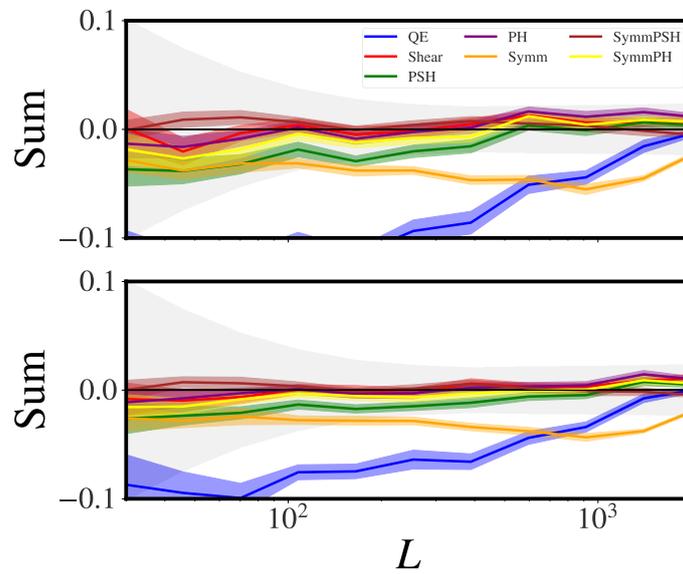
*Fig. B.13* Total relative foreground-induced auto-spectrum CMB lensing bias per mode for each estimator considered in this work, for  $l_{\max} = 3500$ . In the top panel is shown the contribution arising when using the point source mask constructed on 148 GHz data; in the bottom panel, when using the point source mask constructed from the product of point source masks on the SO frequencies. When calculating the ILC combinations we fix the weights to theory ones, without considering the change in mask. We can see how the aggressive masking reduces the biases, with particular emphasis on the QE, meaning that the trispectrum-foreground-induced bias term is the most affected by the masking operation.

that using this expression leads to the same results as not leaving  $w(\vec{L})$  free for some particular cases of  $f_b$ .

## B.7 CMB lensing biases from quadratic estimators

### B.7.1 Biases from simulations

In Figures B.13, B.14 we show foreground biases at ILC for a configuration where CMB modes used in lensing reconstruction come from  $l_{\min} = 30$  and  $l_{\max} = 3500$ , and the amplitude is calculated from CMB lensing scales between  $L_{\min} = 30$  and  $L_{\max} = 1000$ .



*Fig. B.14* Same as Figure B.13 but for the cross-spectrum of CMB lensing with an LSST-like sample. We can see how the aggressive masking reduces the biases, with particular emphasis on the QE.



## Appendices of Chapter 5

### C.1 Derivation of density reconstruction from the bispectrum

We will here consider how we can reconstruct an unknown field  $X$  given the knowledge of i) its bispectrum  $B$  with two other fields  $Y, Z$ , and ii) a measurement of these two other fields. While we write our argument exploiting the connection between bispectra and quadratic estimators in a form that is generally valid, for this paper we will assume that  $X = \delta_1$ , the linear density field, and both  $Y, Z$  are the observed non-linear density field  $Y, Z = \delta_g$ ; given that we can calculate the  $\delta_1 \delta_g \delta_g$  bispectrum, we can easily write an estimator for  $\delta_1$  given an observed  $\delta_g$ . We will assume statistical homogeneity and isotropy of the fields and the bispectrum, which is a good approximation for large-scale structure surveys (although it may be broken for other applications).

We begin by defining the bispectrum of the unknown field with two observed fields:

$$\langle X(\vec{k}_1)Y(\vec{k}_2)Z(\vec{k}_3) \rangle = (2\pi)^3 \delta_D(\vec{k}_1 + \vec{k}_2 + \vec{k}_3) B_{XYZ}(k_1, k_2, k_3) . \quad (\text{C.1})$$

We will now write an ansatz for recovering the unknown field  $X$  from a quadratic estimator involving  $Y, Z$ :

$$\hat{X}(\vec{K}) = \int_{\vec{q}} g(\vec{q}, \vec{K} - \vec{q}) Y(\vec{q}) Z(\vec{K} - \vec{q}) , \quad (\text{C.2})$$

where we have introduced a function  $g$  which weights these pairs of modes. As we will see, the arguments of  $Y, Z$  assumed here are a consequence of the delta function momentum constraint in the bispectrum.

We now derive this function  $g$ . The function must obviously give an unbiased estimator. In a situation where one may not wish to average over  $Y, Z$  at fixed  $X$  (e.g.

because they are the same fields), we will define unbiasedness by the condition that

$$\langle X(\vec{K}')\hat{X}(\vec{K}) \rangle = (2\pi)^3 \delta_{\text{D}}(\vec{K}' + \vec{K}) P_{XX}(K) . \quad (\text{C.3})$$

For the estimator, this implies that since

$$\begin{aligned} \langle X(\vec{K}')\hat{X}(\vec{K}) \rangle &= \int_{\vec{q}} g(\vec{q}, \vec{K} - \vec{q}) \langle X(\vec{K}')Y(\vec{q})Z(\vec{K} - \vec{q}) \rangle \\ &= (2\pi)^3 \delta_{\text{D}}(\vec{K}' + \vec{K}) \int_{\vec{q}} g(\vec{q}, \vec{K} - \vec{q}) B_{XYZ}(K, q, |\vec{K} - \vec{q}|) , \end{aligned}$$

we have a normalization condition on  $g$

$$I[g] \equiv \int_{\vec{q}} g(\vec{q}, \vec{K} - \vec{q}) \frac{B_{XYZ}(K, q, |\vec{K} - \vec{q}|)}{P_{XX}(K)} = 1 . \quad (\text{C.4})$$

We would also like the estimator  $\hat{X}$  to have as little variance per mode as possible. We will assume that, for the purposes of variance calculation, the fields can be approximated as Gaussian and statistically isotropic. Under these assumptions, the variance  $V[f](\vec{K})$  is given by:

$$\begin{aligned} \hat{X}(\vec{K})\hat{X}(\vec{K}') &= (2\pi)^3 V[g](\vec{K}) \delta_{\text{D}}(\vec{K} + \vec{K}') \\ &= \int_{\vec{q}, \vec{q}'} g(\vec{q}, \vec{K} - \vec{q}) g(\vec{q}', \vec{K}' - \vec{q}') Y(\vec{q})Z(\vec{K} - \vec{q})Y(\vec{q}')Z(\vec{K}' - \vec{q}') \\ &= \int_{\vec{q}, \vec{q}'} g(\vec{q}, \vec{K} - \vec{q}) g(\vec{q}', \vec{K}' - \vec{q}') [Y(\vec{q})Y(\vec{q}')Z(\vec{K} - \vec{q})Z(\vec{K}' - \vec{q}') \\ &\quad + Y(\vec{q})Z(\vec{K}' - \vec{q}')Z(\vec{K} - \vec{q})Y(\vec{q}')] \\ &= \int_{\vec{q}, \vec{q}'} g(\vec{q}, \vec{K} - \vec{q}) g(\vec{q}', \vec{K}' - \vec{q}') [(2\pi)^3 P_{YY}(q) \delta_{\text{D}}(\vec{q} + \vec{q}') (2\pi)^3 P_{ZZ}(\vec{K} - \vec{q}) \delta_{\text{D}}(\vec{K} + \vec{K}' - \vec{q} - \vec{q}') \\ &\quad + (2\pi)^3 P_{YZ}(q) \delta_{\text{D}}(\vec{q} + \vec{K}' - \vec{q}') (2\pi)^3 P_{YZ}(\vec{K} - \vec{q}) \delta_{\text{D}}(\vec{K} - \vec{q} + \vec{q}')] \\ &= (2\pi)^3 \delta_{\text{D}}(\vec{K} + \vec{K}') \int_{\vec{q}} \left[ g(\vec{q}, \vec{K} - \vec{q}) (-\vec{q}, -\vec{K} + \vec{q}) P_{YY}(q) P_{ZZ}(\vec{K} - \vec{q}) \right. \\ &\quad \left. g(\vec{q}, \vec{K} - \vec{q}) g(\vec{q} - \vec{K}, -\vec{q}) P_{YZ}(q) P_{YZ}(\vec{K} - \vec{q}) \right] , \quad (\text{C.5}) \end{aligned}$$

where we have used Wick's Theorem. In the following we will specialize to the case of  $Y = Z$  (which is the relevant case for our application) and can thus consider  $g$  to be symmetric under exchange of its arguments. With the requirement for the reconstructed field to be real, this implies the following expression for the variance as

a functional of  $g$ :

$$V[g](\vec{K}) = 2 \int_{\vec{q}} g^2(\vec{q}, \vec{K} - \vec{q}) P_{YY}(q) P_{YY}(\vec{K} - \vec{q}) \quad . \quad (\text{C.6})$$

We can thus solve for  $g$  by minimizing the variance  $V[g](\vec{K})$  subject to the constraint  $I[g] = 1$ . We can do this by introducing a Lagrange multiplier  $N_{\alpha\alpha}$  and minimizing

$$V[g] - N_{\alpha\alpha} \times I[g] \quad , \quad (\text{C.7})$$

with respect to  $g$ . Minimizing this expression, we obtain

$$g(\vec{q}, \vec{K} - \vec{q}) = N_{\alpha\alpha}(K) \frac{1}{4P_{YY}(q)P_{YY}(\vec{K} - \vec{q})} \frac{B_{XY}(K, q, |\vec{K} - \vec{q}|)}{P_{XX}(K)} \quad , \quad (\text{C.8})$$

where from the constraint equation  $I = 1$  we find

$$N_{\alpha\alpha}(K) = \left[ \int_{\vec{q}} \frac{1}{4P_{YY}(q)P_{YY}(\vec{K} - \vec{q})} \left( \frac{B_{XY}(K, q, |\vec{K} - \vec{q}|)}{P_{XX}(K)} \right)^2 \right]^{-1} \quad . \quad (\text{C.9})$$

Applying this to our choice of fields, i.e., evaluating the  $\langle \delta_1 \delta_g \delta_g \rangle$  bispectrum, we note that we recover a function  $g$  which gives the same expression for the quadratic estimator as used in the main part of our paper. With Eq. (5.15) we have for the bispectrum of a linear mode  $X = \delta_1$  and two galaxy modes (ignoring biases)  $Y = Z = \delta_g$

$$B_{\delta_1, \delta_g, \delta_g}(k_1, k_2, k_3) = 2 \left[ F_{\alpha}(\vec{k}_1, \vec{k}_3) P_{\text{lin}}(k_1) P_{\text{lin}}(k_3) + F_{\alpha}(\vec{k}_1, \vec{k}_2) P_{\text{lin}}(k_1) P_{\text{lin}}(k_2) \right] \quad . \quad (\text{C.10})$$

Plugging this into Eq. (C.9) we get

$$g(\vec{q}, \vec{K} - \vec{q}) = N_{\alpha\alpha}(K) \frac{F_{\alpha}(\vec{K}, -\vec{q}) P_{\text{lin}}(q) + F_{\alpha}(\vec{K}, -\vec{K} + \vec{q}) P_{\text{lin}}(|\vec{K} - \vec{q}|)}{2P_{\text{tot}}(q)P_{\text{tot}}(\vec{K} - \vec{q})} \quad , \quad (\text{C.11})$$

which agrees with Eq. (5.15) in the main text.

## C.2 Noise expressions for quadratic estimator

In this appendix, we derive expressions for the noise power spectrum corresponding to the auto-correlation of the reconstructed field  $\hat{\Delta}_{\alpha}(\vec{K})$  and its cross-correlation with the original input tracer field  $\delta_g(\vec{K})$ . We will see that the noise comes from a combination

of shot noise, due to discrete sampling of the underlying matter field, and cosmic variance.

### C.2.1 Noise for the auto correlation of the reconstructed field

In this appendix, we calculate the covariance of our quadratic estimators, which is defined as

$$\begin{aligned} \langle \hat{\Delta}_\alpha(\vec{K}) \hat{\Delta}_\beta(\vec{K}') \rangle - \langle \hat{\Delta}_\alpha(\vec{K}) \rangle \langle \hat{\Delta}_\beta(\vec{K}') \rangle = \\ \int_{\vec{q}} \int_{\vec{q}'} g_\alpha(\vec{q}, \vec{K} - \vec{q}) g_\beta(\vec{q}', \vec{K}' - \vec{q}') \left( \langle \delta_g(\vec{q}) \delta_g(\vec{K} - \vec{q}) \delta_g(\vec{q}') \delta_g(\vec{K}' - \vec{q}') \rangle \right. \\ \left. - \langle \delta_g(\vec{q}) \delta_g(\vec{K} - \vec{q}) \rangle \langle \delta_g(\vec{q}') \delta_g(\vec{K}' - \vec{q}') \rangle \right). \end{aligned} \quad (\text{C.12})$$

To compute this expression we have to first derive the four point function for the input tracer field. It is possible to derive the shot noise formulae directly in Fourier space, with a discretized version of the tracer field. Alternative derivations of bispectrum shot noise can be found in [136] following [87]. We cross-checked our results with [271, 55].

Let us start by rederiving the stochasticity contributions to the power spectrum, bispectrum and trispectrum. Let us consider a finite number  $N$  of point-like tracers, such as galaxies, at positions  $\vec{x}_i$  in a finite volume  $V$ .<sup>1</sup> Their Fourier space density field is then given as a sum of plane waves

$$\delta_g(\vec{k}) = \frac{1}{\bar{n}} \sum_i \exp[i\vec{k} \cdot \vec{x}_i], \quad (\text{C.13})$$

where  $\bar{n} = N/V$ . The power spectrum of the discrete tracers in the finite volume can then be computed as

$$\begin{aligned} P_g(k) &= \frac{1}{V} \langle \delta_g(\vec{k}) \delta_g(-\vec{k}) \rangle \\ &= \frac{V}{N^2} \sum_{i=j} \langle \exp[i\vec{k}(\vec{x}_i - \vec{x}_j)] \rangle + \frac{V}{N^2} \sum_{i \neq j} \langle \exp[i\vec{k}(\vec{x}_i - \vec{x}_j)] \rangle \\ &= \frac{1}{\bar{n}} + P_{g,\text{cont}}(k). \end{aligned} \quad (\text{C.14})$$

Here, the constant  $1/\bar{n}$  is denoted the shot noise term and we have identified the non-zero lag expectation value with the continuous part of the discrete tracer power

<sup>1</sup>It is useful to recall that all wave vectors in a finite volume are integer multiples of the fundamental wavenumber. The Dirac delta distribution thus becomes a Kronecker delta  $(2\pi)^2 \delta_D(\vec{k}_1 + \vec{k}_2) \rightarrow V \delta_K(\vec{k}_1, \vec{k}_2)$ .

spectrum  $P_{\text{g,cont}}(k)$ . In the local bias model at linear order we have  $P_{\text{g,cont}}(k) = b_1^2 P_{\text{lin}}(k)$ , which becomes  $P_{\text{g,cont}}(k) = [b_{10} + b_{01}/M(k)]^2 P_{\text{lin}}(k)$  in the presence of primordial non-Gaussianities of the local kind. Let us now consider the bispectrum. Following the same steps that led to the power spectrum above, we have to consider the case where all three positions coincide, the case where two positions coincide but are different from the third, and finally the case where all three positions are distinct:

$$\begin{aligned}
B_{\text{g}}(\vec{k}_1, \vec{k}_2) &= \frac{1}{V} \langle \delta_{\text{g}}(\vec{k}_1) \delta_{\text{g}}(\vec{k}_2) \delta_{\text{g}}(-\vec{k}_1 - \vec{k}_2) \rangle \\
&= \frac{V^2}{N^3} \sum_{i=j=l} \langle \exp [i\vec{k}_1(\vec{x}_i - \vec{x}_l) + i\vec{k}_2(\vec{x}_j - \vec{x}_l)] \rangle \\
&\quad + \frac{V^2}{N^3} \sum_{i=l \neq j} \langle \exp [i\vec{k}_1(\vec{x}_i - \vec{x}_l) + i\vec{k}_2(\vec{x}_j - \vec{x}_l)] \rangle + 2 \text{ perm.} \quad (\text{C.15}) \\
&\quad + \frac{V^2}{N^3} \sum_{i \neq j, j \neq l, i \neq l} \langle \exp [i\vec{k}_1(\vec{x}_i - \vec{x}_l) + i\vec{k}_2(\vec{x}_j - \vec{x}_l)] \rangle \\
&= \frac{1}{\bar{n}^2} + \frac{1}{\bar{n}} [P_{\text{g,cont}}(k_1) + 2 \text{ perm.}] + B_{\text{g,cont}}(\vec{k}_1, \vec{k}_2).
\end{aligned}$$

Again, the non-zero lag correlators are identified with the continuous power spectrum and bispectrum of the tracer field. We see that two different stochasticity corrections arise: a  $1/\bar{n}^2$  constant shot noise term and a product of the shot noise and the continuous power spectrum. As above for the power spectrum, in the presence of primordial non-Gaussianity, both of these continuous statistics contain the respective non-Gaussian bias corrections. Note that there is now a coupling between stochasticity and clustering which is enhanced with respect to the pure noise term on large scales.

Let us connect this result to the noise terms introduced in Eq. (5.21), where the relevant contributions are given by

$$\delta_{\text{g}} \supset \epsilon + \epsilon_{\delta} \delta + \epsilon_{\varphi} \varphi. \quad (\text{C.16})$$

The three-point correlator of the noise fields  $\epsilon$  can be associated with the white-noise term in Eq. (C.15)

$$\langle \epsilon(\vec{k}_1) \epsilon(\vec{k}_2) \epsilon(\vec{k}_3) \rangle = (2\pi)^3 \delta_{\text{D}}(\vec{k}_1 + \vec{k}_2 + \vec{k}_3) \frac{1}{\bar{n}^2}. \quad (\text{C.17})$$

The contributions from  $\epsilon_\delta \delta$  and  $\epsilon_\varphi \varphi$  arise by correlating with the linear  $\delta$  and  $\varphi$  fields and the linear noise term  $\epsilon$

$$\begin{aligned} & \langle \epsilon(\vec{k}_1) ([\epsilon_\delta \star \delta] + [\epsilon_\varphi \star \varphi]) (\vec{k}_2) (b_{10} \delta + b_{01} \varphi) (\vec{k}_3) \rangle \\ &= \int_{\vec{q}} \left[ \langle \epsilon(\vec{k}_1) \epsilon_\delta(\vec{q}) \rangle \left( b_{10} \langle \delta(\vec{k}_2 - \vec{q}) \delta(\vec{k}_3) \rangle + b_{01} \langle \varphi(\vec{k}_2 - \vec{q}) \delta(\vec{k}_3) \rangle \right) \right. \\ & \left. + \langle \epsilon(\vec{k}_1) \epsilon_\varphi(\vec{q}) \rangle \left( b_{10} \langle \varphi(\vec{k}_2 - \vec{q}) \delta(\vec{k}_3) \rangle + b_{01} \langle \varphi(\vec{k}_2 - \vec{q}) \varphi(\vec{k}_3) \rangle \right) \right]. \end{aligned} \quad (\text{C.18})$$

Using [233, 79] we have for the noise correlators

$$\langle \epsilon \epsilon_\delta \rangle = (2\pi)^3 \delta_{\text{D}}(\vec{k} + \vec{k}') \frac{b_{10}}{\bar{n}}, \quad \langle \epsilon \epsilon_\varphi \rangle = \frac{b_{01}}{b_{10}} \langle \epsilon \epsilon_\delta \rangle = (2\pi)^3 \delta_{\text{D}}(\vec{k} + \vec{k}') \frac{b_{01}}{\bar{n}}. \quad (\text{C.19})$$

We finally obtain for the mixed contribution to the three-point correlator

$$\begin{aligned} \langle \epsilon(\epsilon_\delta \delta + \epsilon_\varphi \varphi) (b_{10} \delta + b_{01} \varphi) \rangle &= (2\pi)^3 \delta_{\text{D}}(\vec{k}_1 + \vec{k}_2 + \vec{k}_3) \frac{P_{\text{lin}}(k_3)}{\bar{n}} \\ & \times \left[ b_{10} \left( b_{10} + \frac{b_{01}}{M(k_3)} \right) + \frac{b_{01}}{M(k_3)} \left( b_{10} + \frac{b_{01}}{M(k_3)} \right) \right]. \end{aligned} \quad (\text{C.20})$$

In summary, we have

$$B_{\text{g}}(k_1, k_2, k_3) = \frac{1}{\bar{n}^2} + \frac{1}{\bar{n}} \left[ \left( b_{10} + \frac{b_{01}}{M(k_1)} \right)^2 P_{\text{lin}}(k_1) + 2 \text{ perm.} \right] + B_{\text{g,cont}}(k_1, k_2, k_3). \quad (\text{C.21})$$

This is equivalent to Eq. (C.15), as long as the galaxy power spectrum  $P_{\text{g,cont}}$  in that equation is taken to be the one with the scale dependent non-Gaussian bias  $b_{10} + b_{01}/M(k)$ .

For the connected trispectrum, we have four positions, which allow for five different configurations: all positions equal, three positions equal but different from the fourth one, two pairs of positions equal but different from the other pair, one pair of positions equal but different from all other positions and finally, all four positions distinct. The

trispectrum can then be written as

$$\begin{aligned}
T_{\text{g,conn}}(\vec{k}_1, \vec{k}_2, \vec{k}_3) &= \frac{1}{V} \left\langle \delta_{\text{g}}(\vec{k}_1) \delta_{\text{g}}(\vec{k}_2) \delta_{\text{g}}(\vec{k}_3) \delta_{\text{g}}(-\vec{k}_1 - \vec{k}_2 - \vec{k}_3) \right\rangle \\
&= \frac{V^3}{N^3} \frac{1}{N} \sum_{i,j,s,t} \left\langle \exp \left[ i\vec{k}_1(\vec{x}_i - \vec{x}_t) + i\vec{k}_2(\vec{x}_j - \vec{x}_t) + i\vec{k}_3(\vec{x}_s - \vec{x}_t) \right] \right\rangle \\
&= \frac{V^3}{N^3} \frac{1}{N} \sum_{i=j=s=t} \left\langle \exp \left[ i\vec{k}_1(\vec{x}_i - \vec{x}_t) + i\vec{k}_2(\vec{x}_j - \vec{x}_t) + i\vec{k}_3(\vec{x}_s - \vec{x}_t) \right] \right\rangle \\
&\quad + \frac{V^2}{N^2} \frac{V}{N^2} \sum_{i=j=s \neq t} \left\langle \exp \left[ -i\vec{k}_4(\vec{x}_i - \vec{x}_t) \right] \right\rangle + 3 \text{ perm.} \\
&\quad + \frac{V^2}{N^2} \frac{V}{N^2} \sum_{i=j \neq s=t} \left\langle \exp \left[ i(\vec{k}_1 + \vec{k}_2)(\vec{x}_i - \vec{x}_t) \right] \right\rangle + 2 \text{ perm.} \\
&\quad + \frac{V}{N} \frac{V^2}{N^3} \sum_{i=j \neq s \neq t, j \neq t} \left\langle \exp \left[ i(\vec{k}_1 + \vec{k}_2)(\vec{x}_i - \vec{x}_t) + i\vec{k}_3(\vec{x}_s - \vec{x}_t) \right] \right\rangle + 5 \text{ perm.} \\
&\quad + \frac{V^3}{N^4} \sum_{i \neq j \neq s \neq t, i \neq t, i \neq s, j \neq t} \left\langle \exp \left[ i\vec{k}_1(\vec{x}_i - \vec{x}_t) + i\vec{k}_2(\vec{x}_j - \vec{x}_t) + i\vec{k}_3(\vec{x}_s - \vec{x}_t) \right] \right\rangle \\
&= \frac{1}{\bar{n}^3} + \frac{1}{\bar{n}^2} [P_{\text{g,cont}}(\vec{k}_4) + 3 \text{ perm.}] + \frac{1}{\bar{n}^2} [P_{\text{g,cont}}(\vec{k}_1 + \vec{k}_2) + 2 \text{ perm.}] \\
&\quad + \frac{1}{\bar{n}} [B_{\text{g,cont}}(\vec{k}_1 + \vec{k}_2, \vec{k}_3) + 5 \text{ perm.}] + T_{\text{g,cont}}(\vec{k}_1, \vec{k}_2, \vec{k}_3).
\end{aligned} \tag{C.22}$$

Furthermore, there is a disconnected cosmic variance contribution for counter-aligned pairs of momenta

$$\begin{aligned}
T_{\text{g,disconn}}(\vec{k}_1, -\vec{k}_1, \vec{k}_3) &= \frac{1}{V} \left\langle \delta_{\text{g}}(\vec{k}_1) \delta_{\text{g}}(-\vec{k}_1) \delta_{\text{g}}(\vec{k}_3) \delta_{\text{g}}(-\vec{k}_3) \right\rangle \\
&= \frac{V^3}{N^3} \frac{1}{N} \sum_{i,j,s,t} \left\langle \exp \left[ i\vec{k}_1(\vec{x}_i - \vec{x}_j) + i\vec{k}_3(\vec{x}_s - \vec{x}_t) \right] \right\rangle \\
&= V \frac{V}{N^2} \sum_{i,j} \left\langle \exp \left[ i\vec{k}_1(\vec{x}_i - \vec{x}_j) \right] \right\rangle \frac{V}{N^2} \sum_{s,t} \left\langle \exp \left[ i\vec{k}_3(\vec{x}_s - \vec{x}_t) \right] \right\rangle.
\end{aligned} \tag{C.23}$$

In the continuous case this disconnected contribution to the four-point function becomes

$$\begin{aligned}
\left\langle \delta_{\text{g}}(\vec{k}_1) \delta_{\text{g}}(\vec{k}_2) \delta_{\text{g}}(\vec{k}_3) \delta_{\text{g}}(\vec{k}_4) \right\rangle &\supset (2\pi)^6 \delta_{\text{D}}(\vec{k}_1 + \vec{k}_2) \delta_{\text{D}}(\vec{k}_3 + \vec{k}_4) \left[ P_{\text{g,cont}}(k_1) + \frac{1}{\bar{n}} \right] \left[ P_{\text{g,cont}}(k_3) + \frac{1}{\bar{n}} \right] \\
&\quad + 2 \text{ perm.} \tag{C.24}
\end{aligned}$$

This will give the Gaussian contribution to the covariance of the estimator.

Thus, the total covariance for the auto-spectrum of the reconstructed field is

$$\begin{aligned}
\langle \hat{\Delta}_\alpha(\vec{K}) \hat{\Delta}_\beta(\vec{K}') \rangle - \langle \hat{\Delta}_\alpha(\vec{K}) \rangle \langle \hat{\Delta}_\beta(\vec{K}') \rangle &= \int_{\vec{q}} \int_{\vec{q}'} g_\alpha(\vec{q}, \vec{K} - \vec{q}) g_\beta(\vec{q}', \vec{K}' - \vec{q}') \\
&\quad \times \left[ \langle \delta_g(\vec{q}) \delta_g(\vec{K} - \vec{q}) \delta_g(\vec{q}') \delta_g(\vec{K}' - \vec{q}') \rangle \right. \\
&\quad \left. - \langle \delta_g(\vec{q}) \delta_g(\vec{K} - \vec{q}) \rangle \langle \delta_g(\vec{q}') \delta_g(\vec{K}' - \vec{q}') \rangle \right] \\
&= (2\pi)^3 \delta_D(\vec{K} + \vec{K}') \left[ \frac{N_{\alpha\alpha}(\vec{K}) N_{\beta\beta}(\vec{K})}{N_{\alpha\beta}(\vec{K})} + N_{\alpha\beta, \text{shot}}(\vec{K}) \right], \tag{C.25}
\end{aligned}$$

where

$$\begin{aligned}
N_{\alpha\beta, \text{shot}}^T(\vec{K}) &\equiv \int_{\vec{q}} \int_{\vec{q}'} g_\alpha(\vec{q}, \vec{K} - \vec{q}) g_\beta(\vec{q}', -\vec{K} - \vec{q}') T_{\text{g,conn}}(\vec{q}, \vec{K} - \vec{q}, \vec{q}', -\vec{K} - \vec{q}') \\
&= \int_{\vec{q}} \int_{\vec{q}'} g_\alpha(\vec{q}, \vec{K} - \vec{q}) g_\beta(\vec{q}', -\vec{K} - \vec{q}') \\
&\quad \times \left\{ \frac{1}{\bar{n}^3} + \frac{1}{\bar{n}^2} \left[ P_{\text{g,cont}}(\vec{q}) + P_{\text{g,cont}}(\vec{K} - \vec{q}) + P_{\text{g,cont}}(\vec{q}') + P_{\text{g,cont}}(\vec{K} + \vec{q}') \right] \right. \\
&\quad + \frac{1}{\bar{n}^2} \left[ P_{\text{g,cont}}(\vec{K}) + P_{\text{g,cont}}(\vec{q} - \vec{K} - \vec{q}') + P_{\text{g,cont}}(\vec{q} + \vec{q}') \right] \\
&\quad + \frac{1}{\bar{n}} \left[ B_{\text{g,cont}}(\vec{K}, \vec{q}', -\vec{K} - \vec{q}') + B_{\text{g,cont}}(\vec{q} + \vec{q}', \vec{K} - \vec{q}, -\vec{K} - \vec{q}') \right. \\
&\quad + B_{\text{g,cont}}(\vec{q} - \vec{K} - \vec{q}', \vec{K} - \vec{q}, \vec{q}') + B_{\text{g,cont}}(\vec{K} - \vec{q} + \vec{q}', \vec{q}, -\vec{K} - \vec{q}') \\
&\quad \left. \left. + B_{\text{g,cont}}(\vec{K} - \vec{q} - \vec{K} - \vec{q}', \vec{q}, \vec{q}') + B_{\text{g,cont}}(\vec{q}' - \vec{K} - \vec{q}', \vec{q}, \vec{K} - \vec{q}) \right] \right\}. \tag{C.26}
\end{aligned}$$

For our forecasts, we use the ‘‘growth’’ estimator, and therefore the conversion to the notation of the main text is

$$P_{\text{rr,shot}}(K) = N_{\text{GG,shot}}^T(K). \tag{C.27}$$

Note an important feature of Eq. (C.26): the shot noise contribution to the quadratic estimator’s noise power spectrum depends on the tracer power spectrum and bispectrum at the same scale  $K$  as the mode being reconstructed. This is to be contrasted with the Gaussian estimator noise in Eq. (5.16), which is mainly determined by the tracer power spectrum at the smallest scale  $k_{\text{max}}$  used in the estimator. Since the tracer power spectrum and bispectrum both increase at smaller wavenumbers (down

to the matter-radiation equality scale), Eq. (C.26)'s sensitivity to large scales can cause it to dominate over the Gaussian estimator noise if  $\bar{n}$  is sufficiently low. In our forecasts in the main text, this condition is met for DESI (Fig. 5.9) and MegaMapper (Fig. 5.11).

We can simplify the above expression by noticing that changes of variables can make some terms of the integrand equivalent. For example, we can simplify the final 3 lines of Eq. (C.26) into

$$\int_{\vec{q}} \int_{\vec{q}'} g_{\alpha}(\vec{q}, \vec{K} - \vec{q}) g_{\beta}(\vec{q}', -\vec{K} - \vec{q}') \frac{1}{\bar{n}} \left( B_{\text{g,cont}}(\vec{K}, \vec{q}', -\vec{K} - \vec{q}') + 4B_{\text{g,cont}}(-\vec{q} - \vec{q}', \vec{q}, \vec{q}') \right. \\ \left. + B_{\text{g,cont}}(-\vec{K}, \vec{q}, \vec{K} - \vec{q}) \right). \quad (\text{C.28})$$

In our calculations, we take the tree-level expression for the bispectrum, obtainable from Eq. (5.37) as (see also [23, 274])

$$B_{\text{g,cont}}(\vec{k}_1, \vec{k}_2, \vec{k}_3) = 2 \left( b_{10} + f_{\text{NL}} \frac{c_{01}}{M(\vec{k}_1)} \right) \left( b_{10} + f_{\text{NL}} \frac{c_{01}}{M(\vec{k}_2)} \right) \sum_{\alpha} c_{\alpha} F_{\alpha}(\vec{k}_1, \vec{k}_2) P_{\text{lin}}(\vec{k}_1) P_{\text{lin}}(\vec{k}_2) + \\ + 2 \text{ perms.} \quad (\text{C.29})$$

### C.2.2 Noise for the cross correlation of the reconstructed field with the tracer field

The noise calculation for the cross-correlation of the reconstructed field with the tracer field is very similar to the one above. The variance of the cross-correlation of the reconstructed field with the tracer field depends on the bispectrum of the tracer field derived above in Eq. (C.15) (see also [136]):

$$\left\langle \hat{\Delta}_{\alpha}(\vec{K}) \delta_{\text{g}}(\vec{K}') \right\rangle - \left\langle \hat{\Delta}_{\alpha}(\vec{K}) \right\rangle \left\langle \delta_{\text{g}}(\vec{K}') \right\rangle = \\ \left\langle \hat{\Delta}_{\alpha}(\vec{K}) \delta_{\text{g}}(\vec{K}') \right\rangle = \\ \int_{\vec{q}} g_{\alpha}(\vec{q}, \vec{K} - \vec{q}) \left\langle \delta_{\text{g}}(\vec{q}) \delta_{\text{g}}(\vec{K} - \vec{q}) \delta_{\text{g}}(\vec{K}') \right\rangle \\ = (2\pi)^3 \delta_{\text{D}}(\vec{K} + \vec{K}') \int_{\vec{q}} g_{\alpha}(\vec{q}, \vec{K} - \vec{q}) \left[ B_{\text{g,cont}}(\vec{q}, \vec{K} - \vec{q}, -\vec{K}) + \right. \\ \left. + \frac{1}{\bar{n}} \left( P_{\text{g,cont}}(\vec{q}) + P_{\text{g,cont}}(\vec{K} - \vec{q}) + P_{\text{g,cont}}(-\vec{K}) \right) + \frac{1}{\bar{n}^2} \right]. \quad (\text{C.30})$$

Then, the shot noise power is given by

$$N_{\alpha,\text{shot}}^{\text{B}}(\vec{K}) \equiv \frac{1}{\bar{n}} \int_{\vec{q}} g_{\alpha}(\vec{q}, \vec{K} - \vec{q}) \left[ P_{\text{g,cont}}(\vec{q}) + P_{\text{g,cont}}(\vec{K} - \vec{q}) + P_{\text{g,cont}}(-\vec{K}) + \frac{1}{\bar{n}} \right]. \quad (\text{C.31})$$

For the growth estimator we use in our forecasts we thus have

$$P_{\text{gr,shot}}(K) = N_{\text{G,shot}}^{\text{B}}(K). \quad (\text{C.32})$$

## C.3 Bias-hardening

In Sec. 5.2.1, we saw that a quadratic estimator designed to have unit response to a specific form of mode-coupling will generically acquire a mean-field contamination from other forms of mode coupling that are not incorporated in the estimator's weights (see Eq. 5.18). Our main approach in this paper is to include that contamination in our model for the estimator's output, marginalizing over the associated free (bias) parameters where necessary. Alternatively, one can attempt to define an estimator that is orthogonal to those extra mode-couplings; such a "bias-hardening" procedure has been applied to weak lensing of the CMB (e.g. [197, 201]) and line intensity maps [91]. In this appendix, we explore this approach and explain why we did not find it to be useful for this study.

### C.3.1 General derivation

First, we derive a form of bias-hardening that is a light generalization of the standard form (e.g. [197]). Recall that a quadratic estimator with weights  $g_{\alpha}$  is given by

$$\hat{\Delta}_{\alpha}(\vec{K}) = \int_{\vec{q}} g_{\alpha}(\vec{q}, \vec{K} - \vec{q}) \delta_{\text{g}}(\vec{q}) \delta_{\text{g}}(\vec{K} - \vec{q}), \quad (\text{C.33})$$

with expectation value

$$\langle \hat{\Delta}_{\alpha}(\vec{K}) \rangle_{\delta_1(\vec{K}) \text{ fixed}} = \sum_{\beta} \left[ c_{\beta} \int_{\vec{q}} g_{\alpha}(\vec{q}, \vec{K} - \vec{q}) f_{\beta}(\vec{q}, \vec{K} - \vec{q}) \right] b_1 \delta_1(\vec{K}). \quad (\text{C.34})$$

The estimator will be unbiased with respect to the  $\alpha$  mode-coupling if

$$\int_{\vec{q}} g_{\alpha}(\vec{q}, \vec{K} - \vec{q}) f_{\alpha}(\vec{q}, \vec{K} - \vec{q}) = 1. \quad (\text{C.35})$$

The  $\beta \neq \alpha$  terms in the sum in Eq. (C.34) could be subtracted if we knew the values of the  $c_\beta$  coefficients ahead of time, but this will generally not be true. Instead, we can attempt to set the weights  $g_\alpha$  such that

$$\int_{\vec{q}} g_\alpha(\vec{q}, \vec{K} - \vec{q}) f_\beta(\vec{q}, \vec{K} - \vec{q}) = 0 , \quad (\text{C.36})$$

for all  $\beta \neq \alpha$ .

We begin by assuming that there is only one additional mode-coupling  $\beta$  that we are concerned with, and requiring that the estimator's response to it is not necessarily zero, but a chosen constant  $c$  instead:

$$\int_{\vec{q}} g_\alpha(\vec{q}, \vec{K} - \vec{q}) f_\beta(\vec{q}, \vec{K} - \vec{q}) = c , \quad (\text{C.37})$$

while also imposing Eq. (C.35) and minimizing the Gaussian contribution to the variance of  $\hat{\Delta}_\alpha(\vec{K})$ , given by

$$\text{Var}_G[\hat{\Delta}_\alpha(\vec{K})] = 2 \int_{\vec{q}} g_\alpha(\vec{q}, \vec{K} - \vec{q}) g_\alpha^*(\vec{q}, \vec{K} - \vec{q}) P_{\text{tot}}(q) P_{\text{tot}}(|\vec{K} - \vec{q}|) . \quad (\text{C.38})$$

We find  $g_\alpha$  that satisfies these conditions by the method of Lagrange multipliers, starting with the following function:

$$\begin{aligned} L[g_\alpha, \lambda, \lambda_*] = & 2 \int_{\vec{q}} g_\alpha(\vec{q}, \vec{K} - \vec{q}) g_\alpha^*(\vec{q}, \vec{K} - \vec{q}) P_{\text{tot}}(\vec{q}) P_{\text{tot}}(\vec{K} - \vec{q}) \\ & - \lambda \left[ \int_{\vec{q}} g_\alpha(\vec{q}, \vec{K} - \vec{q}) f_\alpha(\vec{q}, \vec{K} - \vec{q}) - 1 \right] - \lambda_* \left[ \int_{\vec{q}} g_\alpha(\vec{q}, \vec{K} - \vec{q}) f_\beta(\vec{q}, \vec{K} - \vec{q}) - c \right] . \end{aligned} \quad (\text{C.39})$$

If we wanted to control the effect of other mode-couplings in Eq. (C.34), we would simply add other terms with similar constraints to this equation. We demand that the functional derivative of  $L$  with respect to  $g_\alpha(\vec{q}', \vec{K} - \vec{q}')$  vanishes:

$$\begin{aligned} 0 \stackrel{!}{=} \frac{\delta L[g_\alpha, \lambda, \lambda_*]}{\delta g_\alpha(\vec{q}', \vec{K} - \vec{q}')} &= 2 \int_{\vec{q}} \left[ \delta_D(\vec{q} - \vec{q}') g_\alpha^*(\vec{q}, \vec{K} - \vec{q}) + g_\alpha(\vec{q}, \vec{K} - \vec{q}) \delta_D(\vec{q} - \vec{q}') \right] P_{\text{tot}}(\vec{q}) P_{\text{tot}}(\vec{K} - \vec{q}) \\ &\quad - \lambda \int_{\vec{q}} \delta_D(\vec{q} - \vec{q}') f_\alpha(\vec{q}, \vec{K} - \vec{q}) - \lambda_* \int_{\vec{q}} \delta_D(\vec{q} - \vec{q}') f_\beta(\vec{q}, \vec{K} - \vec{q}) \\ &= 4g_\alpha(\vec{q}', \vec{K} - \vec{q}') P_{\text{tot}}(\vec{q}') P_{\text{tot}}(\vec{K} - \vec{q}') - \lambda f_\alpha(\vec{q}', \vec{K} - \vec{q}') - \lambda_* f_\beta(\vec{q}', \vec{K} - \vec{q}') , \end{aligned} \quad (\text{C.40})$$

where in the last line, we took  $g_\alpha$  to be real. This implies that (relabelling  $\vec{q}' \rightarrow \vec{q}$ )

$$g_\alpha(\vec{q}, \vec{K} - \vec{q}) = \frac{\lambda f_\alpha(\vec{q}, \vec{K} - \vec{q}) + \lambda_* f_\beta(\vec{q}, \vec{K} - \vec{q})}{4P_{\text{tot}}(\vec{q})P_{\text{tot}}(\vec{K} - \vec{q})}, \quad (\text{C.41})$$

and plugging this into Eq. (C.35) gives

$$\lambda = 2N_{\alpha\alpha} - \lambda_* \frac{N_{\alpha\alpha}}{N_{\alpha\beta}}, \quad (\text{C.42})$$

using the definition of  $N_{\alpha\beta}$  from Eq. (5.16). (Note that we obtain the original filter if  $\lambda_* = 0$ .) Inserting Eqs. (C.41) and (C.42) into (C.37), we get<sup>2</sup>

$$\lambda_* = 2 \frac{1}{1 - r_{\alpha\beta}^2} N_{\beta\beta}(\vec{K}) \left[ c(\vec{K}) - \frac{N_{\alpha\alpha}(\vec{K})}{N_{\alpha\beta}(\vec{K})} \right], \quad (\text{C.43})$$

where

$$r_{\alpha\beta}^2 \equiv \frac{N_{\alpha\alpha}N_{\beta\beta}}{N_{\alpha\beta}^2}. \quad (\text{C.44})$$

Thus, the final form of the weight function is

$$\begin{aligned} g_\alpha(\vec{q}, \vec{K} - \vec{q}) &= \frac{N_{\alpha\alpha}(\vec{K})f_\alpha(\vec{q}, \vec{K} - \vec{q})}{2P_{\text{tot}}(\vec{q})P_{\text{tot}}(\vec{K} - \vec{q})} \\ &\quad + \frac{f_{\beta_*}(\vec{q}, \vec{K} - \vec{q})N_{\beta\beta}(\vec{K}) - \frac{N_{\alpha\alpha}(\vec{K})N_{\beta\beta}(\vec{K})}{N_{\alpha\beta}(\vec{K})}f_\alpha(\vec{q}, \vec{K} - \vec{q})}{2P_{\text{tot}}(\vec{q})P_{\text{tot}}(\vec{K} - \vec{q})} \\ &\times \frac{1}{1 - r_{\alpha\beta}^2} \left[ c(\vec{K}) - \frac{N_{\alpha\alpha}(\vec{K})}{N_{\alpha\beta_*}(\vec{K})} \right]. \end{aligned} \quad (\text{C.45})$$

This is the standard filter for mode-coupling  $\alpha$ , plus some additional terms related to the response to mode-coupling  $\beta$ . For the case  $c = N_{\alpha\alpha}N_{\alpha\beta}^{-1}$ , one obtains the standard estimator and contamination term.

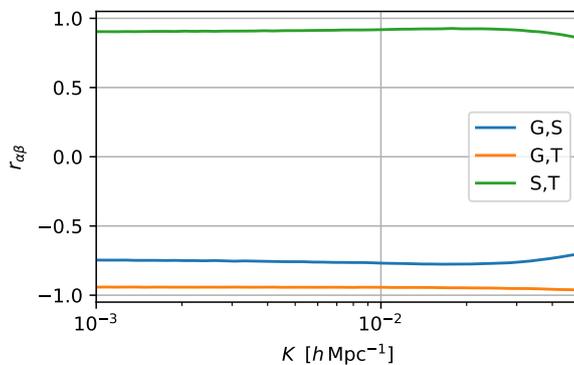
<sup>2</sup>To generalise to several mode-couplings, the coefficients  $\lambda_{i*}$  are obtained from  $\vec{\lambda}_* = -2\mathbf{A}^{-1}\vec{c}$ , where  $\mathbf{A}$  has elements  $A_{ij} = (N_{\beta_i\beta_j}^{-1} - N_{\alpha\alpha}N_{\alpha\beta_j}^{-1}N_{\alpha\beta_i}^{-1})$  and  $\vec{c}$  has elements  $N_{\alpha\alpha}N_{\alpha\beta_i}^{-1}$ .

### C.3.2 Application to long-mode reconstruction

To relate the bias-hardened estimator derived above to the specific application we consider in this paper, let us examine the variance of the estimator:

$$\text{Var}[\hat{\Delta}_\alpha^{\text{H}}(\vec{K})] = N_{\alpha\alpha}(\vec{K}) + N_{\beta\beta}(\vec{K}) \frac{\left[c(\vec{K}) - \frac{N_{\alpha\alpha}(\vec{K})}{N_{\alpha\beta}(\vec{K})}\right]^2}{1 - r_{\alpha\beta}^2(\vec{K})}. \quad (\text{C.46})$$

If  $|c(\vec{K})| < |N_{\alpha\alpha}(\vec{K})N_{\alpha\beta}(\vec{K})^{-1}|$ , then the increase in the estimator's variance scales with  $(1 - r_{\alpha\beta}^2)^{-1}$ , where  $r_{\alpha\beta}$  is the correlation coefficient between the un-hardened estimator  $\hat{\Delta}_\alpha$  and the analogous estimator for the other mode-coupling,  $\hat{\Delta}_\beta$ . In Fig. C.1, we show  $r_{\alpha\beta}(K)$  for the growth, shift, and tidal mode-couplings introduced in Sec. 5.2.1. It is clear that the corresponding estimators are highly correlated, so that any bias-hardened estimator will have a much larger variance than without bias-hardening. In our numerical tests (with  $c = 0$ ), when one of these three estimators was bias-hardened with respect to the other two, we found that the variance increased enough to eliminate any advantages of removing the mean-field contamination, and therefore we did not implement any bias-hardening in our final forecasts.



*Fig. C.1* Cross-correlation coefficients between G, S, and T estimators, as given by Eq. (C.44), for our DESI-like survey forecast (the other surveys give similar results). We see that all three estimators are highly correlated, implying, through Eq. (C.46), that applying bias-hardening will strongly increase the noise of the resulting estimator.

In the course of this investigation, we derived a compact form for a bias-hardened quadratic estimator in the case of three mode-couplings, and we reproduce this derivation here in case it may be useful in other contexts. Considering only G, S, and T, the expectation values of the corresponding quadratic estimators (see Eq. 5.18) can

be written in matrix form:

$$\begin{bmatrix} \langle \hat{\Delta}_G \rangle \\ \langle \hat{\Delta}_S \rangle \\ \langle \hat{\Delta}_T \rangle \end{bmatrix} = b_1 \begin{bmatrix} 1 & N_{GG}N_{GS}^{-1} & N_{GG}N_{GT}^{-1} \\ N_{SS}N_{SG}^{-1} & 1 & N_{SS}N_{TS}^{-1} \\ N_{TT}N_{TG}^{-1} & N_{TT}N_{TS}^{-1} & 1 \end{bmatrix} \begin{bmatrix} c_G \\ c_S \\ c_T \end{bmatrix} \delta_1 . \quad (\text{C.47})$$

We can derive bias-hardened estimators (in the  $c(\vec{K}) = 0$  case) by inverting this system, solving for each  $b_1 c_\alpha \delta_1$ . After some lengthy algebra, the results can be written in terms of the original variances plus certain combinations of the original cross-correlation coefficients:

$$\begin{aligned} \text{Var}[\hat{\Delta}_G^H] &= N_{GG} \times \frac{1 - r_{ST}^2}{\det M} , \\ \text{Var}[\hat{\Delta}_S^H] &= N_{SS} \times \frac{1 - r_{GT}^2}{\det M} , \\ \text{Var}[\hat{\Delta}_T^H] &= N_{TT} \times \frac{1 - r_{GS}^2}{\det M} , \end{aligned} \quad (\text{C.48})$$

where  $M$  denotes the matrix in Eq. (C.47), and

$$\det M = 1 - r_{GS}^2 - r_{GT}^2 - r_{ST}^2 + 2r_{GS}r_{GT}r_{ST} . \quad (\text{C.49})$$

There can be nontrivial cancellations within the above determinant, and in our case, these lead to large increases in the variances of the bias-hardened estimators.

Finally, we mention a few other possible solutions to the problem of mean-field contamination. Instead of fixing the contamination to some value, as in Eq. (C.37), one could require it to be smaller than some fixed value, or one could minimise some total function that depends on the contamination and the (Gaussian) variance. For example it is possible to define a ‘‘bouncing estimator’’ by solving the following minimization problem:

$$\begin{aligned} L[g_\alpha, \lambda, \vec{K}] &= 2 \int_{\vec{q}} |g_\alpha(\vec{q}, \vec{K} - \vec{q})|^2 P_{\text{tot}}(\vec{q}) P_{\text{tot}}(\vec{K} - \vec{q}) + A \left[ \int_{\vec{q}} g_\alpha(\vec{q}, \vec{K} - \vec{q}) f_\beta(\vec{q}, \vec{K} - \vec{q}) \right]^2 \\ &\quad - \lambda \left[ \int_{\vec{q}} g_\alpha(\vec{q}, \vec{K} - \vec{q}) f_\alpha(\vec{q}, \vec{K} - \vec{q}) - 1 \right] . \end{aligned} \quad (\text{C.50})$$

The intuition behind this is that we want to minimize the variance of the  $\alpha$  estimator, trying also to take into account the contamination from the other mode-coupling. The square is to ensure that the modulus of the contamination is minimized in the combination that makes  $L$  the smallest. If we go to the standard minimum variance solution, the solution here will ‘‘bounce’’ from it, because it would increase  $L$ , if we take into account the square of the contamination. We also want to decrease the

contamination, but without it taking a large negative value. The last term enforces the standard unbiasedness condition.

We reiterate, however, that for some applications, such as the  $f_{\text{NL}}$  constraints we consider in this paper, a mean-field contamination can actually be advantageous. We leave the problem of finding a fully optimal estimator for long-wavelength reconstruction to future work.

## C.4 Contamination of quadratic estimator by $f_{\text{NL}}$ terms: analytical expressions

The overdensity of a biased tracer has second order contributions in the linear field coming from the presence of primordial non-Gaussianity. In this appendix, we show that when reconstructing the new field on large scales, we get an  $f_{\text{NL}}$  term proportional to  $\frac{1}{M(\vec{K})}$ . We will show this for the low- $|\vec{K}|$  limit, which is the relevant regime for the reconstructed modes we are concerned with.

### C.4.1 Expansion of basic quantities

We need to expand a few quantities first. Expanding the linear power spectrum around  $\vec{q}$  gives

$$\begin{aligned}
 \lim_{|\vec{K}| \rightarrow 0} P_{\text{lin}}(|\vec{K} - \vec{q}|) &= P_{\text{lin}}(|\vec{q}|) - \vec{\nabla}_{\vec{q}} P_{\text{lin}} \cdot \vec{K} + O(|\vec{K}|^2) \\
 &= P_{\text{lin}}(|\vec{q}|) - \frac{\partial P_{\text{lin}}}{\partial |\vec{q}|} \vec{\nabla}_{\vec{q}} |\vec{q}| \cdot \vec{K} + O(|\vec{K}|^2) \\
 &= P_{\text{lin}}(\vec{q}) \left( 1 - \frac{|\vec{K}| \mu}{|\vec{q}|} \frac{\partial \ln P_{\text{lin}}}{\partial \ln |\vec{q}|} \right) + O(|\vec{K}|^2) \\
 &\equiv P_{\text{lin}}(\vec{q}) (1 + \Delta_P) + O(|\vec{K}|^2),
 \end{aligned} \tag{C.51}$$

where the last expression is useful because we can see the expansion in powers of  $\frac{|\vec{K}|}{|\vec{q}|}$ , and where  $\mu$  is the cosine of the angle between  $\vec{K}$  and  $\vec{q}$ . In the same way, we can also expand for

$$\lim_{|\vec{K}| \rightarrow 0} P_{\text{NL}}(|\vec{K} - \vec{q}|) = P_{\text{NL}}(\vec{q}) \left( 1 - \frac{|\vec{K}| \mu}{|\vec{q}|} \frac{\partial \ln P_{\text{NL}}}{\partial \ln |\vec{q}|} \right) + O(|\vec{K}|^2) \equiv P_{\text{NL}}(\vec{q}) (1 + \Delta_{\text{NL}}) + O(|\vec{K}|^2), \tag{C.52}$$

and

$$\lim_{|\vec{K}| \rightarrow 0} M|\vec{K} - \vec{q}| = M(\vec{q}) \left( 1 - \frac{|\vec{K}| \mu}{|\vec{q}|} \frac{\partial \ln M}{\partial \ln |\vec{q}|} \right) + O(|\vec{K}|^2) \equiv M(\vec{q}) (1 + \Delta_M). \quad (\text{C.53})$$

We can also write

$$\lim_{|\vec{K}| \rightarrow 0} |\vec{K} - \vec{q}| = \lim_{|\vec{K}| \rightarrow 0} \sqrt{|\vec{K}|^2 + |\vec{q}|^2 - 2|\vec{K}||\vec{q}|\mu} \approx |\vec{q}| \left( 1 - \frac{|\vec{K}|}{|\vec{q}|} \mu \right), \quad (\text{C.54})$$

and the expression for the cosine of the angle between  $\vec{K}$  and  $\vec{K} - \vec{q}$  as

$$\mu' = \lim_{|\vec{K}| \rightarrow 0} \approx -\mu + (1 - \mu^2) \frac{|\vec{K}|}{|\vec{q}|} + \frac{|\vec{K}|^2}{|\vec{q}|^2} \mu \approx -\mu + (1 - \mu^2) \frac{|\vec{K}|}{|\vec{q}|}. \quad (\text{C.55})$$

## C.4.2 Expansion of mode coupling expressions

As  $|\vec{K}| \rightarrow 0$  for the next calculations we will assume  $\frac{|\vec{K}|}{|\vec{q}|} \ll 1$  so that we keep terms linear in this variable. If the linear terms completely cancel, we include terms at the next order.

Recall the definition of  $f_\beta(\vec{q}, \vec{K} - \vec{q})$  from Eq. (5.10):

$$f_\beta(\vec{q}, \vec{K} - \vec{q}) = 2 \left[ F_\alpha(\vec{K}, -\vec{q}) P_{\text{lin}}(|\vec{q}|) + F_\alpha(\vec{K}, -\vec{K} + \vec{q}) P_{\text{lin}}(|\vec{K} - \vec{q}|) \right]. \quad (\text{C.56})$$

Starting from Eq. (C.56), we have that for the  $\varphi\varphi$  term

$$\begin{aligned} f_{\varphi\varphi}(\vec{q}, \vec{K} - \vec{q}) &= \frac{2}{M(\vec{K})} \left[ \frac{M(\vec{K} - \vec{q})}{M(-\vec{q})} P_{\text{lin}}(\vec{q}) + \vec{q} \rightarrow \vec{K} - \vec{q} \right] \\ &= \frac{2P_{\text{lin}}(\vec{q})}{M(\vec{K})} \left[ \frac{M(\vec{q})(1 + \Delta_M(\vec{q}))}{M(\vec{q})} + \frac{M(\vec{q})}{M(\vec{q})(1 + \Delta_M(\vec{q}))} (1 + \Delta_P(\vec{q})) \right] \\ &\approx \frac{2P_{\text{lin}}(\vec{q})}{M(\vec{K})} [2 + \Delta_P(\vec{q})], \end{aligned} \quad (\text{C.57})$$

and for the  $b_{01}$  term

$$\begin{aligned}
 f_{01}(\vec{q}, \vec{K} - \vec{q}) &= \frac{2}{M(\vec{K})} \frac{1}{2} \left[ (\vec{K} \cdot (-\vec{q})) \left( \frac{1}{|\vec{q}|^2} + \frac{M(\vec{K})}{|\vec{K}|^2 M(-\vec{q})} \right) P_{\text{lin}}(-\vec{q}) + \vec{q} \rightarrow \vec{K} - \vec{q} \right] \\
 &= \frac{1}{2} \frac{2P_{\text{lin}}(\vec{q})}{M(\vec{K})} \left[ \vec{K} \cdot (-\vec{q}) \left( \frac{1}{|\vec{q}|^2} + \frac{M(\vec{K})}{|\vec{K}|^2 M(\vec{q})} \right) + \vec{K} \cdot (-1)(\vec{K} - \vec{q}) \right. \\
 &\quad \left. \times \left( \frac{1}{|\vec{q}|^2} (1 + \frac{|\vec{K}|}{|\vec{q}|} \mu) + \frac{M(\vec{K})}{K^2 M(\vec{q})} (1 - \Delta_M(\vec{q})) \right) \times (1 + \Delta_P(\vec{q})) \right] \\
 &\approx \frac{1}{2} \frac{2P_{\text{lin}}(\vec{q})}{M(\vec{K})} \left[ -\frac{|\vec{K}|^2}{|\vec{q}|^2} - \frac{|\vec{K}|^3}{|\vec{q}|^3} \mu - \frac{M(\vec{K})}{M(\vec{q})} (1 + \Delta_P(\vec{q}) - \Delta_M(\vec{q})) + \frac{|\vec{K}|^2}{|\vec{q}|^2} \mu^2 \right. \\
 &\quad \left. + \frac{q\mu M(\vec{K})}{|\vec{K}| M(\vec{q})} (\Delta_P(\vec{q}) - \Delta_M(\vec{q})) \right] \\
 &\approx \frac{1}{2} \frac{2P_{\text{lin}}(\vec{q})}{M(\vec{q})} \left[ -1 - \Delta_P(\vec{q}) - \Delta_M(\vec{q}) - (1 - \mu^2) \frac{|\vec{K}|^2}{|\vec{q}|^2} \right], \quad (\text{C.58})
 \end{aligned}$$

where we remember that  $M(\vec{q}) \propto |\vec{q}|^{-2}$ . For the  $b_{11}$  term,

$$\begin{aligned}
 f_{11}(\vec{q}, \vec{K} - \vec{q}) &= \frac{2}{M(\vec{K})} \frac{1}{2} \left[ \left( 1 + \frac{M(\vec{K})}{M(-\vec{q})} \right) P_{\text{lin}}(-\vec{q}) + \vec{q} \rightarrow \vec{K} - \vec{q} \right] \\
 &= \frac{P_{\text{lin}}(\vec{q})}{M(\vec{K})} \left[ 1 + \frac{M(\vec{K})}{M(\vec{q})} + 1 + \frac{M(\vec{K})}{M(\vec{q})} (1 - \Delta_M(\vec{q})) (1 + \Delta_P(\vec{q})) \right] \quad (\text{C.59}) \\
 &\approx \frac{P_{\text{lin}}(\vec{q})}{M(\vec{K})} \left[ 2 + \frac{M(\vec{K})}{M(\vec{q})} (1 + \Delta_P(\vec{q}) - \Delta_M(\vec{q})) \right],
 \end{aligned}$$

and finally for the  $b_{02}$  term

$$\begin{aligned}
 f_{02}(\vec{q}, \vec{K} - \vec{q}) &= \frac{2}{M(\vec{K})} \left[ \frac{P_{\text{lin}}(-\vec{q})}{M(-\vec{q})} + \vec{q} \rightarrow \vec{K} - \vec{q} \right] \\
 &= \frac{2P_{\text{lin}}(\vec{q})}{M(\vec{K})} \left[ \frac{1}{M(\vec{q})} + \frac{1}{M(\vec{q})} (1 - \Delta_M(\vec{q}) + \Delta_P(\vec{q})) \right] \quad (\text{C.60}) \\
 &\approx \frac{2P_{\text{lin}}(\vec{q})}{M(\vec{K})} \frac{1}{M(\vec{q})} \left[ 2 - \Delta_M(\vec{q}) + \Delta_P(\vec{q}) \right].
 \end{aligned}$$

We wrote all of them in such a way that, when possible, we can factor out a  $\frac{1}{M(\vec{K})}$ . Finally the growth term can be written as

$$f_{\text{G}}(\vec{q}, \vec{K} - \vec{q}) = 2 \left[ \frac{17}{21} P(\vec{q}) + \frac{17}{21} P_{\text{lin}}(\vec{K} - \vec{q}) \right] = 2 \frac{17}{21} P(\vec{q}) \left[ 2 + \Delta_{\text{P}}(\vec{q}) \right]. \quad (\text{C.61})$$

### C.4.3 Writing the large scale contamination terms

Recall that the terms that contaminate the expectation value of the quadratic estimator are of the form  $c_{\alpha} \frac{N_{\text{GG}}}{N_{\text{G}\alpha}}$  (see Eq. 5.18). Therefore, for the large scale limit, we need to consider

$$\begin{aligned} \lim_{|\vec{K}| \rightarrow 0} N_{\text{G}\alpha}^{-1}(\vec{K}) &= \lim_{|\vec{K}| \rightarrow 0} \frac{2\pi}{(2\pi)^3} \int_{-1}^1 d\mu \int_{q_{\text{min}}}^{q_{\text{max}}} dq q^2 \frac{f_{\alpha}(\vec{q}, \vec{K} - \vec{q})}{2P_{\text{tot}}(\vec{q})P_{\text{tot}}(\vec{K} - \vec{q})} f_{\text{g}}(\vec{q}, \vec{K} - \vec{q}) \\ &= \lim_{|\vec{K}| \rightarrow 0} \frac{1}{b_1^4} \frac{2\pi}{(2\pi)^3} \int_{-1}^1 d\mu \int_{q_{\text{min}}}^{q_{\text{max}}} dq q^2 \frac{f_{\alpha}(\vec{q}, \vec{K} - \vec{q})}{2P_{\text{NL}}^2(\vec{q})} 2 \frac{17}{21} P_{\text{lin}}(\vec{q}) [2 + \Delta_{\text{P}}(\vec{q})] (1 - \Delta_{\text{PN}}), \end{aligned} \quad (\text{C.62})$$

where we assume no shot noise in the total galaxy power spectrum. Dropping any non-zero power of  $\frac{|\vec{K}|}{|\vec{q}|}$ , for small  $\vec{K}$  (with respect to reconstruction modes  $\vec{q}$ ), we obtain

$$\begin{aligned} N_{\text{GG}}(\vec{K}) &\approx \left[ \frac{1}{b_1^4} \frac{2\pi}{(2\pi)^3} \left( 2 \frac{17}{21} \right)^2 4 \int_{q_{\text{min}}}^{q_{\text{max}}} dq q^2 \frac{P_{\text{lin}}^2(\vec{q})}{P_{\text{NL}}^2(\vec{q})} [1 + \Delta_{\text{P}}(\vec{q}) - \Delta_{\text{PN}}(\vec{q})] \right]^{-1} \\ &\approx \left[ \frac{1}{b_1^4} \frac{4\pi}{(2\pi)^3} 8 \left( \frac{17}{21} \right)^2 \int_{q_{\text{min}}}^{q_{\text{max}}} dq q^2 \frac{P_{\text{lin}}^2(\vec{q})}{P_{\text{NL}}^2(\vec{q})} \right]^{-1}, \end{aligned} \quad (\text{C.63})$$

and we can see that if we approximate  $P_{\text{lin}} \approx P_{\text{NL}}$  for the small scales of reconstruction, then the Gaussian reconstruction noise is roughly proportional to the volume shell between  $q_{\text{min}}$  and  $q_{\text{max}}$ . Thus, we can take the noise as roughly proportional to  $q_{\text{max}}^3$ , although in practice this relation is not exactly correct.

At this point, we can start listing the  $N_{\text{G}\alpha}$  terms where  $\alpha$  is a mode-coupling involving  $f_{\text{NL}}$ . We begin with  $\alpha = \varphi\varphi$ :

$$N_{\text{G}\varphi\varphi}^{-1}(\vec{K}) \approx \frac{1}{b_1^4} \frac{1}{2\pi^2} \left( \frac{136}{21} \right) \frac{1}{M(\vec{K})} \int_{q_{\text{min}}}^{q_{\text{max}}} dq q^2 \frac{P_{\text{lin}}^2(\vec{q})}{P_{\text{NL}}^2(\vec{q})}, \quad (\text{C.64})$$

such that multiplying by Eq. (C.63), we have a scaling for the  $\alpha = \varphi\varphi$  term in the bias of the reconstructed field:

$$N_{\text{GG}}(\vec{K}) N_{\text{G}\varphi\varphi}^{-1}(\vec{K}) \approx \left( \frac{21}{17} \right) \frac{1}{M(\vec{K})}. \quad (\text{C.65})$$

Similarly we can calculate an approximate expression for the  $\alpha = 01$  term:

$$N_{\text{G}01}^{-1}(\vec{K}) \approx (-1) \frac{1}{b_1^4} \frac{1}{2\pi^2} \frac{1}{2} \frac{68}{21} \int_{q_{\text{min}}}^{q_{\text{max}}} dq q^2 \frac{P_{\text{lin}}^2(\vec{q})}{P_{\text{NL}}^2(\vec{q})} \frac{1}{M(\vec{q})}. \quad (\text{C.66})$$

Multiplying this by the approximate  $N_{\text{GG}}$  expression, we obtain

$$N_{\text{GG}}(\vec{K}) N_{\text{G}01}^{-1}(\vec{K}) \propto -\frac{42}{17}, \quad (\text{C.67})$$

implying that on large scales for the  $\alpha = 01$  term we do not have a  $\frac{1}{K^2}$  behaviour, but a negative constant bias.

Turning to the  $\alpha = 02$  and  $11$  terms, we find

$$N_{\text{G}02}^{-1}(\vec{K}) \approx \frac{1}{b^4} \frac{1}{2\pi^2} \left( \frac{136}{21} \right) \frac{1}{M(\vec{K})} \int_{q_{\text{min}}}^{q_{\text{max}}} dq q^2 \frac{P_{\text{lin}}^2(\vec{q})}{P_{\text{NL}}^2(\vec{q})} \frac{1}{M(\vec{q})}, \quad (\text{C.68})$$

and

$$N_{\text{G}11}^{-1}(\vec{K}) \approx \frac{1}{b^4} \frac{1}{2\pi^2} \left( \frac{68}{21} \right) \frac{1}{M(\vec{K})} \int_{q_{\text{min}}}^{q_{\text{max}}} dq q^2 \frac{P_{\text{lin}}^2(\vec{q})}{P_{\text{NL}}^2(\vec{q})} \left( 1 + \frac{M(\vec{K})}{2M(\vec{q})} \right). \quad (\text{C.69})$$

In both cases we end up with a  $1/K^2$  behavior:

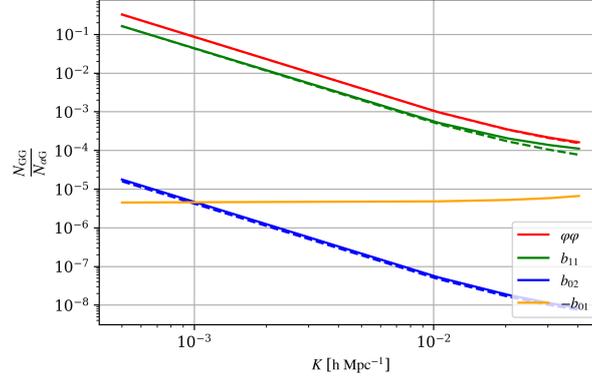
$$N_{\text{GG}}(\vec{K}) N_{\text{G}02}^{-1}(\vec{K}) = 21/17 \frac{1}{M(\vec{K})} \left( \int_{q_{\text{min}}}^{q_{\text{max}}} dq q^2 \frac{P_{\text{lin}}^2(\vec{q})}{P_{\text{NL}}^2(\vec{q})} \frac{1}{M(\vec{q})} / \int_{q_{\text{min}}}^{q_{\text{max}}} dq q^2 \frac{P_{\text{lin}}^2(\vec{q})}{P_{\text{NL}}^2(\vec{q})} \right), \quad (\text{C.70})$$

$$N_{\text{GG}}(\vec{K}) N_{\text{G}11}^{-1}(\vec{K}) = \frac{1}{2} \frac{21}{17} \frac{1}{M(\vec{K})}. \quad (\text{C.71})$$

To summarise, we have found an induced contamination on the G estimator of the following form:

$$\sum_{\alpha \in \{\varphi\varphi, 01, 11, 02\}} \int_{\vec{q}} c_{\alpha} g_{\text{G}}(\vec{q}, \vec{K} - \vec{q}) f_{\alpha}(\vec{q}, \vec{K} - \vec{q}) = \frac{1}{M(|\vec{K}|)} \left[ f_{\text{NL}} A(\vec{K}) + f_{\text{NL}}^2 B(\vec{K}) \right], \quad (\text{C.72})$$

where  $A, B$  are some functions that can be calculated from the definitions or numerically. In Fig. C.2, we show that the analytical approximations are in excellent agreement with the full numerical computations for the contamination curves.



*Fig. C.2* We plot the contamination curves from numerical and analytic approximation. It can be seen that for the  $\varphi\varphi$ ,  $b_{11}$  and  $b_{02}$  terms we have a  $\frac{1}{K^2}$  behaviour. For the analytical calculation of the  $b_{02}$  term we use an approximation  $\left( \int_{q_{\min}}^{q_{\max}} dq q^2 \frac{P_{\text{lin}}^2(\vec{q})}{P_{\text{NL}}^2(\vec{q})} \frac{1}{M(\vec{q})} / \int_{q_{\min}}^{q_{\max}} dq q^2 \frac{P_{\text{lin}}^2(\vec{q})}{P_{\text{NL}}^2(\vec{q})} \right) \approx \frac{1}{2M(q_{\min})}$ . We also show the absolute value of the numerical  $b_{01}$  curve, with an approximate constant value on large scales.

## C.5 Foregrounds for 21 cm intensity mapping

### C.5.1 Implementation in forecasts

As discussed in Sec. 5.4.3, the presence of foregrounds in 21 cm intensity mapping limits the modes of  $\delta_{\text{g}}$  that can be directly observed. Specifically, foregrounds impose a minimum  $k_{\parallel}$  value for these modes, and also obscure modes within a wedge-shaped region in the  $k_{\parallel} - k_{\perp}$  plane. The modes within this wedge satisfy (e.g. [19])

$$k_{\parallel} < \beta(z)k_{\perp} , \quad (\text{C.73})$$

where

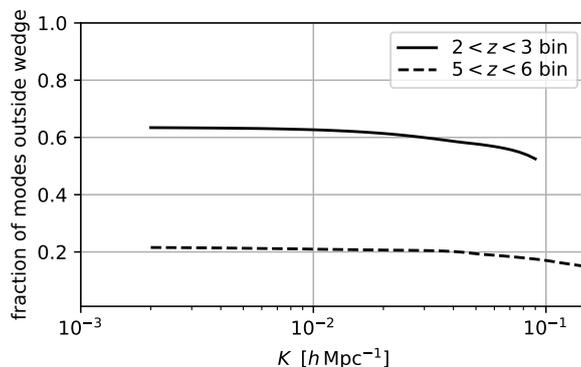
$$\beta(z) \equiv \frac{\chi(z)H(z)}{c(1+z)} \sin(\theta_{\text{w}}) , \quad (\text{C.74})$$

$\chi(z)$  is the comoving distance to redshift  $z$ , and  $\theta_{\text{w}}$  is the maximum angle from the beam center at which the power of a spectrally-smooth source will leak into other regions of Fourier space. The angle  $\theta_{\text{w}}$  is typically related to the width of the primary beam; following [19], we take it to be 3 times the primary beam width of PUMA, or  $\theta_{\text{w}} \approx 3 \times 1.22\lambda(z)/D_{\text{eff}}$ , where  $\lambda(z) = 21(1+z)$  cm and  $D_{\text{eff}} \approx 5$  m is the effective dish diameter ( $\eta_{\text{a}}^{1/2} \times 6$  m with aperture efficiency factor  $\eta_{\text{a}} = 0.7$ ). Using the mean redshifts of the low- $z$  and high- $z$  bins we use in our forecasts, this yields  $\beta \approx 0.38$  and 1.3 for each bin respectively.

The wedge will restrict the small-scale modes that can be used for reconstructing the longer modes via the quadratic estimator in Eq. (5.17), and we can account for this by restricting the reconstruction noise integral in Eq. (5.16) to modes outside the wedge. This means that, when reconstructing a mode with wavevector  $\vec{K}$ , the integration variable  $\vec{q}$  must satisfy

$$|q_{\parallel}| > \beta q_{\perp} \ , \quad |K_{\parallel} - q_{\parallel}| > \beta \left( \vec{K} - \vec{q} \right)_{\perp} \ . \quad (\text{C.75})$$

Rather than implementing these restrictions directly in the integral for  $N_{\alpha\beta}$ , which would cause the result to depend on the full vector  $\vec{K}$  instead of just the norm  $K$ , we use an approximation based on the fact that in the  $q \gg K$  limit,  $N_{\alpha\beta}$  scales like the inverse of the number of modes that contribute to the reconstruction.<sup>3</sup> Thus, the effect of the wedge is mostly to rescale  $N_{\alpha\beta}$  by the inverse of the fraction of modes that are outside of the wedge, i.e. the fraction of the integration domain that satisfies Eq. (C.75). This fraction will depend slightly on the direction of  $\vec{K}$ , and we account for this dependence by averaging the fraction over  $\mu_K \equiv K_{\parallel}/K$ , although the dependence is only mild.



*Fig. C.3* The fraction of small-scale modes lying outside the foreground wedge, for the low- $z$  (*solid*) and high- $z$  (*dashed*) PUMA redshift bins we use in our forecasts. To a good approximation, the reconstruction noise integrals  $N_{\alpha\beta}$  will simply be scaled by the inverse of these fractions in the presence of the wedge.

We plot this angle-averaged fraction in Fig. C.3 for both redshift bins we use for our PUMA forecast. In the  $2 < z < 3$  bin, we find that around 60% of the small-scale modes are untouched by the wedge, while for the  $5 < z < 6$  bin, only 20% of the modes remain, corresponding to a factor of 5 increase in the reconstruction noise compared to the no-wedge case.

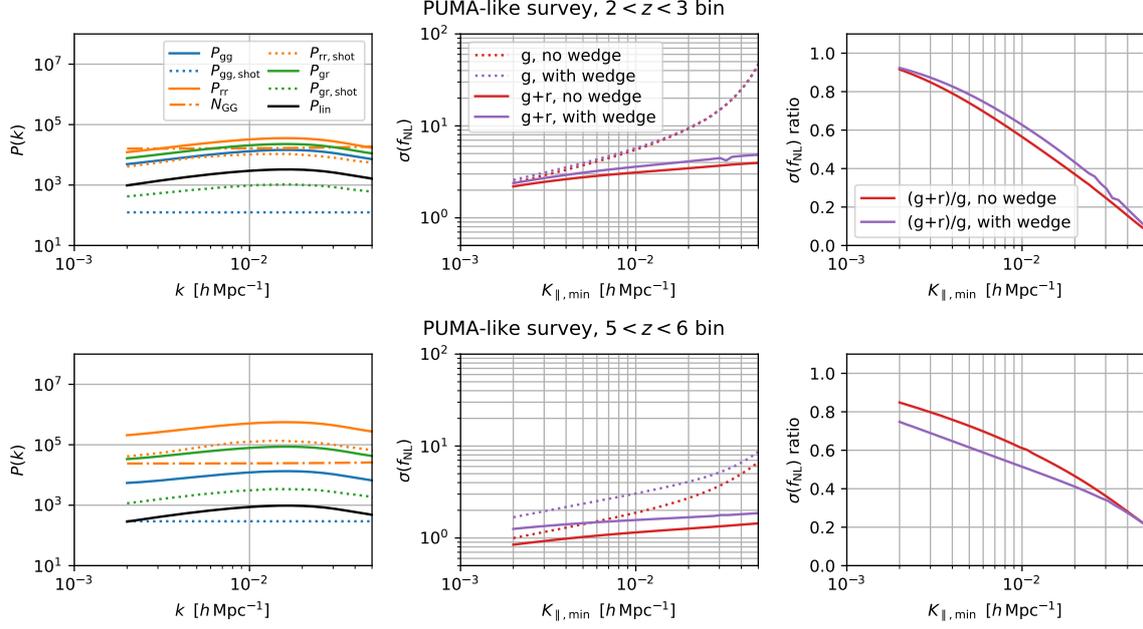
<sup>3</sup>This is identical to what happens to the noise on the standard quadratic estimator for CMB lensing in the  $\ell \gg L$  limit (e.g. [110]).

For the shot noise contributions to the  $\delta_r$  auto spectrum and  $\delta_r$ - $\delta_g$  cross spectrum (Eqs. C.26-C.27 and C.31-C.32, respectively), we directly implement the wedge in the angular limits of the integrals, but we find that it has a negligible effect, since these integrals are normalized with  $N_{\alpha\beta}$  and the fractional change in the integrals and  $N_{\alpha\beta}$  is very similar.

We also need to incorporate the wedge when we integrate the Fisher matrix in Eq. (5.43) over long-wavelength modes used for the  $f_{\text{NL}}$  constraint (Eq. 5.44). Within the wedge, we will not have access to  $\delta_g$ , but we will have access to modes  $\delta_r$  reconstructed with the quadratic estimator. Thus, as in our baseline forecasts with an isotropic  $K_{\text{min}}$ , we sum the outside-wedge and inside-wedge Fisher matrices, each restricted to the appropriate integration domain, with the latter Fisher matrix determined solely by the covariance of the reconstructed modes.

In the next subsection, where we consider a  $K_{\parallel,\text{min}}$  instead of an isotropic  $K_{\text{min}}$ , we likewise implement the restriction  $K > K_{\parallel,\text{min}}$  in Eq. (5.44), and add the contribution from reconstructed modes with  $K < K_{\parallel,\text{min}}$ .

### C.5.2 PUMA forecasts with $K_{\parallel, \min}$



*Fig. C.4* Forecasts for a PUMA-like survey, analogous to Fig. 5.13 except for using a cutoff on the line-of-sight component of accessible  $\delta_g$  modes instead of an isotropic  $K$  cutoff. This is motivated by the fact that 21 cm foregrounds will preferentially obscure modes with low wavenumber components along the line of sight. The absolute values of  $\sigma(f_{\text{NL}})$  are slightly higher when using  $K_{\parallel, \min}$  instead of  $K_{\min}$ , since more modes are eliminated with a  $K_{\parallel}$  cut, but the improvement in  $\sigma(f_{\text{NL}})$  from including reconstructed modes is qualitatively similar to the case with  $K_{\min}$ .

In Sec. 5.4.5, we showed forecasts assuming an isotropic  $K_{\min}$  for  $\delta_g$ . In Fig. C.4, we repeat those forecasts, but with a  $K_{\parallel, \min}$ , assuming that all values of  $K_{\perp}$  within the survey volume can be accessed. The absolute values we find for  $\sigma(f_{\text{NL}})$  are slightly higher, due to the number of inaccessible modes being larger with a  $K_{\parallel, \min}$  cutoff, but the results for the improvement in  $\sigma(f_{\text{NL}})$  due to the inclusion of reconstructed modes are qualitatively similar to those in Sec. 5.4.5.

## References

- [1] Muntazir Mehdi Abidi and Tobias Baldauf. “Cubic Halo Bias in Eulerian and Lagrangian Space”. In: *JCAP* 07 (2018), p. 029. DOI: [10.1088/1475-7516/2018/07/029](https://doi.org/10.1088/1475-7516/2018/07/029). arXiv: [1802.07622](https://arxiv.org/abs/1802.07622) [[astro-ph.CO](#)].
- [2] Y. Sultan Abylkairov et al. “Partially constrained internal linear combination: A method for low-noise CMB foreground mitigation”. In: *Physical Review D* 103.10 (2021). ISSN: 2470-0029. DOI: [10.1103/physrevd.103.103510](https://doi.org/10.1103/physrevd.103.103510). URL: <http://dx.doi.org/10.1103/PhysRevD.103.103510>.
- [3] P. A. R. Ade et al. “Planck2015 results”. In: *Astronomy Astrophysics* 594 (2016), A15. ISSN: 1432-0746. DOI: [10.1051/0004-6361/201525941](https://doi.org/10.1051/0004-6361/201525941). URL: <http://dx.doi.org/10.1051/0004-6361/201525941>.
- [4] P.A.R. Ade et al. “Planck 2015 results. XIII. Cosmological parameters”. In: *Astron. Astrophys.* 594 (2016), A13. DOI: [10.1051/0004-6361/201525830](https://doi.org/10.1051/0004-6361/201525830). arXiv: [1502.01589](https://arxiv.org/abs/1502.01589) [[astro-ph.CO](#)].
- [5] Peter Ade et al. “The Simons Observatory: science goals and forecasts”. In: 2019.2, 056 (Feb. 2019), p. 056. DOI: [10.1088/1475-7516/2019/02/056](https://doi.org/10.1088/1475-7516/2019/02/056). arXiv: [1808.07445](https://arxiv.org/abs/1808.07445) [[astro-ph.CO](#)].
- [6] Saroj Adhikari, Donghui Jeong, and Sarah Shandera. “Constraining primordial and gravitational mode coupling with the position-dependent bispectrum of the large-scale structure”. In: *Phys. Rev. D* 94.8 (2016), p. 083528. DOI: [10.1103/PhysRevD.94.083528](https://doi.org/10.1103/PhysRevD.94.083528). arXiv: [1608.05139](https://arxiv.org/abs/1608.05139) [[astro-ph.CO](#)].
- [7] Amir Aghamousa et al. “The DESI Experiment Part I: Science, Targeting, and Survey Design”. In: (2016). arXiv: [1611.00036](https://arxiv.org/abs/1611.00036) [[astro-ph.IM](#)].
- [8] N. Aghanim et al. “Planck 2018 results”. In: *Astronomy Astrophysics* 641 (2020), A6. ISSN: 1432-0746. DOI: [10.1051/0004-6361/201833910](https://doi.org/10.1051/0004-6361/201833910). URL: <http://dx.doi.org/10.1051/0004-6361/201833910>.
- [9] Nabila Aghanim, Subhabrata Majumdar, and Joseph Silk. “Secondary anisotropies of the CMB”. In: *Reports on Progress in Physics* 71.6 (2008), p. 066902. ISSN: 1361-6633. DOI: [10.1088/0034-4885/71/6/066902](https://doi.org/10.1088/0034-4885/71/6/066902). URL: <http://dx.doi.org/10.1088/0034-4885/71/6/066902>.

- [10] Simone Aiola et al. “The Atacama Cosmology Telescope: DR4 Maps and Cosmological Parameters”. In: *arXiv e-prints*, arXiv:2007.07288 (July 2020), arXiv:2007.07288. arXiv: [2007.07288](#) [[astro-ph.CO](#)].
- [11] Y. Akrami et al. “Planck 2018 results. IX. Constraints on primordial non-Gaussianity”. In: (May 2019). arXiv: [1905.05697](#) [[astro-ph.CO](#)].
- [12] Y. Akrami et al. “Planck 2018 results. X. Constraints on inflation”. In: (July 2018). arXiv: [1807.06211](#) [[astro-ph.CO](#)].
- [13] Y. Akrami et al. “Planck2018 results”. In: *Astronomy Astrophysics* 641 (2020), A10. ISSN: 1432-0746. DOI: [10.1051/0004-6361/201833887](#). URL: <http://dx.doi.org/10.1051/0004-6361/201833887>.
- [14] Y. Akrami et al. “Planck2018 results”. In: *Astronomy & Astrophysics* 641 (2020), A4. ISSN: 1432-0746. DOI: [10.1051/0004-6361/201833881](#). URL: <http://dx.doi.org/10.1051/0004-6361/201833881>.
- [15] Shadab Alam et al. “The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological analysis of the DR12 galaxy sample”. In: 470.3 (2017), pp. 2617–2652. DOI: [10.1093/mnras/stx721](#). arXiv: [1607.03155](#) [[astro-ph.CO](#)].
- [16] Rupert Allison et al. “The Atacama Cosmology Telescope: measuring radio galaxy bias through cross-correlation with lensing”. In: 451.1 (2015), pp. 849–858. DOI: [10.1093/mnras/stv991](#). arXiv: [1502.06456](#) [[astro-ph.CO](#)].
- [17] Marcelo Alvarez et al. “Testing Inflation with Large Scale Structure: Connecting Hopes with Reality”. In: (2014). arXiv: [1412.4671](#) [[astro-ph.CO](#)].
- [18] Stefania Amodeo et al. “Atacama Cosmology Telescope: Modeling the gas thermodynamics in BOSS CMASS galaxies from kinematic and thermal Sunyaev-Zel’dovich measurements”. In: *Phys. Rev. D* 103 (6 2021), p. 063514. DOI: [10.1103/PhysRevD.103.063514](#). URL: <https://link.aps.org/doi/10.1103/PhysRevD.103.063514>.
- [19] Réza Ansari et al. “Inflation and Early Dark Energy with a Stage II Hydrogen Intensity Mapping experiment”. In: (2018). arXiv: [1810.09572](#) [[astro-ph.CO](#)].
- [20] Astropy Collaboration. “Astropy: A community Python package for astronomy”. In: 558, A33 (Oct. 2013), A33. DOI: [10.1051/0004-6361/201322068](#). arXiv: [1307.6212](#) [[astro-ph.IM](#)].
- [21] Tobias Baldauf. *Advanced cosmology part III lecture notes*. 2021.
- [22] Tobias Baldauf. “Modelling large scale structure statistics for precision cosmology”. PhD thesis. Zürich: University of Zurich, 2013. URL: <https://doi.org/10.5167/uzh-164224>.
- [23] Tobias Baldauf, Uros Seljak, and Leonardo Senatore. “Primordial non-Gaussianity in the Bispectrum of the Halo Density Field”. In: *JCAP* 04 (2011), p. 006. DOI: [10.1088/1475-7516/2011/04/006](#). arXiv: [1011.1513](#) [[astro-ph.CO](#)].
- [24] Tobias Baldauf et al. “Evidence for Quadratic Tidal Tensor Bias from the Halo Bispectrum”. In: *Phys. Rev. D* 86 (2012), p. 083540. DOI: [10.1103/PhysRevD.86.083540](#). arXiv: [1201.4827](#) [[astro-ph.CO](#)].
- [25] Tobias Baldauf et al. “Halo stochasticity from exclusion and nonlinear clustering”. In: *Phys. Rev. D* 88.8 (2013), p. 083507. DOI: [10.1103/PhysRevD.88.083507](#). arXiv: [1305.2917](#) [[astro-ph.CO](#)].

- [26] Mario Ballardini, William Luke Matthewson, and Roy Maartens. “Constraining primordial non-Gaussianity using two galaxy surveys and CMB lensing”. In: *Mon. Not. Roy. Astron. Soc.* 489.2 (2019), pp. 1950–1956. DOI: [10.1093/mnras/stz2258](https://doi.org/10.1093/mnras/stz2258). arXiv: [1906.04730](https://arxiv.org/abs/1906.04730) [[astro-ph.CO](#)].
- [27] J. M. Bardeen et al. “The statistics of peaks of Gaussian random fields”. In: *The Astrophysical Journal* 304 (1986), p. 15. ISSN: 0004-637X. DOI: [10.1086/164143](https://doi.org/10.1086/164143). URL: <http://adsabs.harvard.edu/doi/10.1086/164143>.
- [28] Alexandre Barreira and Fabian Schmidt. “Response Approach to the Matter Power Spectrum Covariance”. In: *JCAP* 1711.11 (2017), p. 051. DOI: [10.1088/1475-7516/2017/11/051](https://doi.org/10.1088/1475-7516/2017/11/051). arXiv: [1705.01092](https://arxiv.org/abs/1705.01092) [[astro-ph.CO](#)].
- [29] Alexandre Barreira and Fabian Schmidt. “Responses in Large-Scale Structure”. In: *JCAP* 1706.06 (2017), p. 053. DOI: [10.1088/1475-7516/2017/06/053](https://doi.org/10.1088/1475-7516/2017/06/053). arXiv: [1703.09212](https://arxiv.org/abs/1703.09212) [[astro-ph.CO](#)].
- [30] Nicola Bartolo. *Cosmology Notes*.
- [31] Daniel Baumann. *Cosmology (part III lecture notes)*.
- [32] Daniel Baumann. *TASI Lectures on Inflation*. 2012. arXiv: [0907.5424](https://arxiv.org/abs/0907.5424) [[hep-th](#)].
- [33] Daniel Baumann et al. “Cosmological Non-Linearities as an Effective Fluid”. In: (Apr. 2010). DOI: [10.1088/1475-7516/2012/07/051](https://doi.org/10.1088/1475-7516/2012/07/051). URL: <http://arxiv.org/abs/1004.2488><http://dx.doi.org/10.1088/1475-7516/2012/07/051>.
- [34] E. J. Baxter et al. “A measurement of CMB cluster lensing with SPT and DES year 1 data”. In: 476.2 (2018), pp. 2674–2688. DOI: [10.1093/mnras/sty305](https://doi.org/10.1093/mnras/sty305). arXiv: [1708.01360](https://arxiv.org/abs/1708.01360) [[astro-ph.CO](#)].
- [35] E. J. Baxter et al. “A Measurement of Gravitational Lensing of the Cosmic Microwave Background by Galaxy Clusters Using Data from the South Pole Telescope”. In: 806.2, 247 (2015), p. 247. DOI: [10.1088/0004-637X/806/2/247](https://doi.org/10.1088/0004-637X/806/2/247). arXiv: [1412.7521](https://arxiv.org/abs/1412.7521) [[astro-ph.CO](#)].
- [36] E. J. Baxter et al. “Dark Energy Survey Year 1 results: Methodology and projections for joint analysis of galaxy clustering, galaxy lensing, and CMB lensing two-point functions”. In: 99.2, 023508 (2019), p. 023508. DOI: [10.1103/PhysRevD.99.023508](https://doi.org/10.1103/PhysRevD.99.023508). arXiv: [1802.05257](https://arxiv.org/abs/1802.05257) [[astro-ph.CO](#)].
- [37] Lucia Fonseca de la Bella et al. “The matter power spectrum in redshift space using effective field theory”. In: *JCAP* 1711 (2017), p. 039. DOI: [10.1088/1475-7516/2017/11/039](https://doi.org/10.1088/1475-7516/2017/11/039). arXiv: [1704.05309](https://arxiv.org/abs/1704.05309) [[astro-ph.CO](#)].
- [38] F Bernardeau et al. “Large-scale structure of the Universe and cosmological perturbation theory”. In: *Physrep* 367.December 2001 (2002), pp. 1–3. ISSN: 03701573. DOI: [10.1016/S0370-1573\(02\)00135-7](https://doi.org/10.1016/S0370-1573(02)00135-7). URL: [http://adsabs.harvard.edu/cgi-bin/nph-bib\\_query?bibcode=2002PhR...367....1B&db\\_key=AST](http://adsabs.harvard.edu/cgi-bin/nph-bib_query?bibcode=2002PhR...367....1B&db_key=AST).
- [39] Edmund Bertschinger. “Cosmological Dynamics”. In: (). URL: <https://arxiv.org/abs/astro-ph/9503125>.
- [40] Matteo Biagetti et al. “Verifying the consistency relation for the scale-dependent bias from local primordial non-Gaussianity”. In: *Mon. Not. Roy. Astron. Soc.* 468.3 (2017), pp. 3277–3288. DOI: [10.1093/mnras/stx714](https://doi.org/10.1093/mnras/stx714). arXiv: [1611.04901](https://arxiv.org/abs/1611.04901) [[astro-ph.CO](#)].

- [41] Michael Blanton et al. “The Physical Origin of Scale-dependent Bias in Cosmological Simulations”. In: *The Astrophysical Journal* 522.2 (1999), pp. 590–603. DOI: [10.1086/307660](https://doi.org/10.1086/307660). URL: <https://doi.org/10.1086%2F307660>.
- [42] L. E. Bleem et al. “A Measurement of the Correlation of Galaxy Surveys with CMB Lensing Convergence Maps from the South Pole Telescope”. In: 753.1, L9 (July 2012), p. L9. DOI: [10.1088/2041-8205/753/1/L9](https://doi.org/10.1088/2041-8205/753/1/L9). arXiv: [1203.4808](https://arxiv.org/abs/1203.4808) [[astro-ph.CO](https://arxiv.org/abs/1203.4808)].
- [43] L. E. Bleem et al. *CMB/kSZ and Compton-y Maps from 2500 square degrees of SPT-SZ and Planck Survey Data*. 2021. arXiv: [2102.05033](https://arxiv.org/abs/2102.05033) [[astro-ph.CO](https://arxiv.org/abs/2102.05033)].
- [44] Martin Bucher and Thibaut Louis. “Filling in cosmic microwave background map missing data using constrained Gaussian realizations”. In: 424.3 (2012), pp. 1694–1713. DOI: [10.1111/j.1365-2966.2012.21138.x](https://doi.org/10.1111/j.1365-2966.2012.21138.x). arXiv: [1109.0286](https://arxiv.org/abs/1109.0286) [[astro-ph.CO](https://arxiv.org/abs/1109.0286)].
- [45] Martin Bucher and Thibaut Louis. “Filling in cosmic microwave background map missing data using constrained Gaussian realizations”. In: *Monthly Notices of the Royal Astronomical Society* 424.3 (2012), 1694–1713. ISSN: 0035-8711. DOI: [10.1111/j.1365-2966.2012.21138.x](https://doi.org/10.1111/j.1365-2966.2012.21138.x). URL: <http://dx.doi.org/10.1111/j.1365-2966.2012.21138.x>.
- [46] Joyce Byun et al. “Towards optimal cosmological parameter recovery from compressed bispectrum statistics”. In: *Mon. Not. Roy. Astron. Soc.* 471.2 (2017), pp. 1581–1618. DOI: [10.1093/mnras/stx1681](https://doi.org/10.1093/mnras/stx1681). arXiv: [1705.04392](https://arxiv.org/abs/1705.04392) [[astro-ph.CO](https://arxiv.org/abs/1705.04392)].
- [47] Paolo Cabella and Marc Kamionkowski. “Theory of cosmic microwave background polarization”. In: *International School of Gravitation and Cosmology: The Polarization of the Cosmic Microwave Background*. Mar. 2004. arXiv: [astro-ph/0403392](https://arxiv.org/abs/astro-ph/0403392).
- [48] John Joseph M Carrasco, Mark P. Hertzberg, and Leonardo Senatore. “The effective field theory of cosmological large scale structures”. In: *Journal of High Energy Physics* 2012.9 (2012), pp. 1–41. ISSN: 11266708. DOI: [10.1007/JHEP09\(2012\)082](https://doi.org/10.1007/JHEP09(2012)082).
- [49] John Joseph M. Carrasco et al. “The Effective Field Theory of Large Scale Structures at Two Loops”. In: *JCAP* 07 (2014), p. 057. DOI: [10.1088/1475-7516/2014/07/057](https://doi.org/10.1088/1475-7516/2014/07/057). arXiv: [1310.0464](https://arxiv.org/abs/1310.0464) [[astro-ph.CO](https://arxiv.org/abs/1310.0464)].
- [50] Sean Carroll. *Spacetime and geometry: an introduction to general relativity; International ed.* Essex: Pearson Education, 2013. DOI: [1292026634](https://doi.org/10.1093/acprof:oso/9780195389409). URL: <https://cds.cern.ch/record/1602292>.
- [51] Emanuele Castorina and Azadeh Moradinezhad Dizgah. “Local Primordial Non-Gaussianities and Super-Sample Variance”. In: (May 2020). arXiv: [2005.14677](https://arxiv.org/abs/2005.14677) [[astro-ph.CO](https://arxiv.org/abs/2005.14677)].
- [52] Emanuele Castorina et al. “Redshift-weighted constraints on primordial non-Gaussianity from the clustering of the eBOSS DR14 quasars in Fourier space”. In: *JCAP* 09 (2019), p. 010. DOI: [10.1088/1475-7516/2019/09/010](https://doi.org/10.1088/1475-7516/2019/09/010). arXiv: [1904.08859](https://arxiv.org/abs/1904.08859) [[astro-ph.CO](https://arxiv.org/abs/1904.08859)].
- [53] Marco Celoria and Sabino Matarrese. *Primordial Non-Gaussianity*. 2018. arXiv: [1812.08197](https://arxiv.org/abs/1812.08197) [[astro-ph.CO](https://arxiv.org/abs/1812.08197)].

- [54] Anthony Challinor and Hiranya Peiris. “Lecture notes on the physics of cosmic microwave background anisotropies”. In: *AIP Conference Proceedings* 1132.1 (2009), pp. 86–140. DOI: [10.1063/1.3151849](https://doi.org/10.1063/1.3151849). eprint: <https://aip.scitation.org/doi/pdf/10.1063/1.3151849>. URL: <https://aip.scitation.org/doi/abs/10.1063/1.3151849>.
- [55] Kwan Chuen Chan and Linda Blot. “Assessment of the Information Content of the Power Spectrum and Bispectrum”. In: *Phys. Rev. D* 96.2 (2017), p. 023528. DOI: [10.1103/PhysRevD.96.023528](https://doi.org/10.1103/PhysRevD.96.023528). arXiv: [1610.06585](https://arxiv.org/abs/1610.06585) [[astro-ph.CO](https://arxiv.org/abs/1610.06585)].
- [56] Shi-Fan Chen, Zvonimir Vlah, and Martin White. “Consistent modeling of velocity statistics and redshift-space distortions in one-loop perturbation theory”. In: *Journal of Cosmology and Astroparticle Physics* 2020.07 (2020), 062–062. ISSN: 1475-7516. DOI: [10.1088/1475-7516/2020/07/062](https://doi.org/10.1088/1475-7516/2020/07/062). URL: <http://dx.doi.org/10.1088/1475-7516/2020/07/062>.
- [57] Shi-Fan Chen et al. “Redshift-space distortions in Lagrangian perturbation theory”. In: *Journal of Cosmology and Astroparticle Physics* 2021.03 (2021), p. 100. ISSN: 1475-7516. DOI: [10.1088/1475-7516/2021/03/100](https://doi.org/10.1088/1475-7516/2021/03/100). URL: <http://dx.doi.org/10.1088/1475-7516/2021/03/100>.
- [58] Chi-Ting Chiang. “Halo squeezed-limit bispectrum with primordial non-Gaussianity: A power spectrum response approach”. In: *Phys. Rev. D* 95.12 (2017), p. 123517. DOI: [10.1103/PhysRevD.95.123517](https://doi.org/10.1103/PhysRevD.95.123517). arXiv: [1701.03374](https://arxiv.org/abs/1701.03374) [[astro-ph.CO](https://arxiv.org/abs/1701.03374)].
- [59] Chi-Ting Chiang et al. “Position-dependent correlation function from the SDSS-III Baryon Oscillation Spectroscopic Survey Data Release 10 CMASS Sample”. In: *JCAP* 09 (2015), p. 028. DOI: [10.1088/1475-7516/2015/9/028](https://doi.org/10.1088/1475-7516/2015/9/028). arXiv: [1504.03322](https://arxiv.org/abs/1504.03322) [[astro-ph.CO](https://arxiv.org/abs/1504.03322)].
- [60] Chi-Ting Chiang et al. “Position-dependent power spectrum of the large-scale structure: a novel method to measure the squeezed-limit bispectrum”. In: *JCAP* 05 (2014), p. 048. DOI: [10.1088/1475-7516/2014/05/048](https://doi.org/10.1088/1475-7516/2014/05/048). arXiv: [1403.3411](https://arxiv.org/abs/1403.3411) [[astro-ph.CO](https://arxiv.org/abs/1403.3411)].
- [61] Chi-Ting Chiang et al. “Position-dependent power spectrum of the large-scale structure: a novel method to measure the squeezed-limit bispectrum”. In: *JCAP* 05 (2014), p. 048. DOI: [10.1088/1475-7516/2014/05/048](https://doi.org/10.1088/1475-7516/2014/05/048). arXiv: [1403.3411](https://arxiv.org/abs/1403.3411) [[astro-ph.CO](https://arxiv.org/abs/1403.3411)].
- [62] Steve K. Choi et al. “The Atacama Cosmology Telescope: A Measurement of the Cosmic Microwave Background Power Spectra at 98 and 150 GHz”. In: *arXiv e-prints*, arXiv:2007.07289 (July 2020), arXiv:2007.07289. arXiv: [2007.07289](https://arxiv.org/abs/2007.07289) [[astro-ph.CO](https://arxiv.org/abs/2007.07289)].
- [63] Steve K. Choi et al. “The Atacama Cosmology Telescope: a measurement of the Cosmic Microwave Background power spectra at 98 and 150 GHz”. In: *Journal of Cosmology and Astroparticle Physics* 2020.12 (2020), 045–045. ISSN: 1475-7516. DOI: [10.1088/1475-7516/2020/12/045](https://doi.org/10.1088/1475-7516/2020/12/045). URL: <http://dx.doi.org/10.1088/1475-7516/2020/12/045>.
- [64] CMB-S4 Collaboration. “CMB-S4 Science Book, First Edition”. In: (2016). arXiv: [1610.02743](https://arxiv.org/abs/1610.02743) [[astro-ph.CO](https://arxiv.org/abs/1610.02743)].
- [65] Shaun Cole and Nick Kaiser. “Biased clustering in the cold dark matter cosmogony.” In: 237 (Apr. 1989), pp. 1127–1146. DOI: [10.1093/mnras/237.4.1127](https://doi.org/10.1093/mnras/237.4.1127).

- [66] Asantha Cooray, Marc Kamionkowski, and Robert R. Caldwell. “Cosmic shear of the microwave background: The curl diagnostic”. In: 71.12, 123527 (2005), p. 123527. DOI: [10.1103/PhysRevD.71.123527](https://doi.org/10.1103/PhysRevD.71.123527). arXiv: [astro-ph/0503002](https://arxiv.org/abs/astro-ph/0503002) [[astro-ph](#)].
- [67] Asantha Cooray and Ravi Sheth. “Halo Models of Large Scale Structure”. In: (June 2002). DOI: [10.1016/S0370-1573\(02\)00276-4](https://doi.org/10.1016/S0370-1573(02)00276-4). URL: <http://arxiv.org/abs/astro-ph/0206508>[http://dx.doi.org/10.1016/S0370-1573\(02\)00276-4](http://dx.doi.org/10.1016/S0370-1573(02)00276-4).
- [68] Ji-Ping Dai, Licia Verde, and Jun-Qing Xia. “What Can We Learn by Combining the Skew Spectrum and the Power Spectrum?” In: (2020). arXiv: [2002.09904](https://arxiv.org/abs/2002.09904) [[astro-ph.CO](#)].
- [69] Neal Dalal et al. “The imprints of primordial non-gaussianities on large-scale structure: scale dependent bias and abundance of virialized objects”. In: *Phys. Rev. D* 77 (2008), p. 123514. DOI: [10.1103/PhysRevD.77.123514](https://doi.org/10.1103/PhysRevD.77.123514). arXiv: [0710.4560](https://arxiv.org/abs/0710.4560) [[astro-ph](#)].
- [70] Guido D’Amico et al. “The Cosmological Analysis of the SDSS/BOSS data from the Effective Field Theory of Large-Scale Structure”. In: *JCAP* 05 (2020), p. 005. DOI: [10.1088/1475-7516/2020/05/005](https://doi.org/10.1088/1475-7516/2020/05/005). arXiv: [1909.05271](https://arxiv.org/abs/1909.05271) [[astro-ph.CO](#)].
- [71] Omar Darwish et al. *Density reconstruction from biased tracers and its application to primordial non-Gaussianity*. 2020. arXiv: [2007.08472](https://arxiv.org/abs/2007.08472) [[astro-ph.CO](#)].
- [72] Omar Darwish et al. “The Atacama Cosmology Telescope: a CMB lensing mass map over 2100 square degrees of sky and its cross-correlation with BOSS-CMASS galaxies”. In: *Monthly Notices of the Royal Astronomical Society* 500.2 (2020), 2250â2263. ISSN: 1365-2966. DOI: [10.1093/mnras/staa3438](https://doi.org/10.1093/mnras/staa3438). URL: <http://dx.doi.org/10.1093/mnras/staa3438>.
- [73] Sudeep Das, Josquin Errard, and David Spergel. “Can CMB Lensing Help Cosmic Shear Surveys?” In: *arXiv e-prints*, arXiv:1311.2338 (2013), arXiv:1311.2338. arXiv: [1311.2338](https://arxiv.org/abs/1311.2338) [[astro-ph.CO](#)].
- [74] Sudeep Das and David N. Spergel. “Measuring distance ratios with CMB-galaxy lensing cross-correlations”. In: 79.4, 043509 (2009), p. 043509. DOI: [10.1103/PhysRevD.79.043509](https://doi.org/10.1103/PhysRevD.79.043509). arXiv: [0810.3931](https://arxiv.org/abs/0810.3931) [[astro-ph](#)].
- [75] Sudeep Das et al. “Detection of the Power Spectrum of Cosmic Microwave Background Lensing by the Atacama Cosmology Telescope”. In: 107.2, 021301 (2011), p. 021301. DOI: [10.1103/PhysRevLett.107.021301](https://doi.org/10.1103/PhysRevLett.107.021301). arXiv: [1103.2124](https://arxiv.org/abs/1103.2124) [[astro-ph.CO](#)].
- [76] J. Delabrouille and J. F. Cardoso. “Diffuse Source Separation in CMB Observations”. In: *Data Analysis in Cosmology*. Ed. by V. J. Martínez et al. Vol. 665. 2009, pp. 159–205. DOI: [10.1007/978-3-540-44767-2\\_6](https://doi.org/10.1007/978-3-540-44767-2_6).
- [77] J. Delabrouille et al. “A full sky, low foreground, high resolution CMB map from WMAP”. In: *Astronomy Astrophysics* 493.3 (2008), 835–857. ISSN: 1432-0746. DOI: [10.1051/0004-6361:200810514](https://doi.org/10.1051/0004-6361:200810514). URL: <http://dx.doi.org/10.1051/0004-6361:200810514>.
- [78] V. Desjacques and U. Seljak. “Primordial non-Gaussianity from the large scale structure”. In: *Class. Quant. Grav.* 27 (2010), p. 124011. DOI: [10.1088/0264-9381/27/12/124011](https://doi.org/10.1088/0264-9381/27/12/124011). arXiv: [1003.5020](https://arxiv.org/abs/1003.5020) [[astro-ph.CO](#)].

- [79] Vincent Desjacques, Donghui Jeong, and Fabian Schmidt. “Large-Scale Galaxy Bias”. In: *Phys. Rept.* 733 (2018), pp. 1–193. DOI: [10.1016/j.physrep.2017.12.002](https://doi.org/10.1016/j.physrep.2017.12.002). arXiv: [1611.09787](https://arxiv.org/abs/1611.09787) [[astro-ph.CO](#)].
- [80] Vincent Desjacques, Donghui Jeong, and Fabian Schmidt. “The Galaxy Power Spectrum and Bispectrum in Redshift Space”. In: *JCAP* 12 (2018), p. 035. DOI: [10.1088/1475-7516/2018/12/035](https://doi.org/10.1088/1475-7516/2018/12/035). arXiv: [1806.04015](https://arxiv.org/abs/1806.04015) [[astro-ph.CO](#)].
- [81] Scott Dodelson and Fabian Schmidt. *Modern cosmology*. Elsevier, AP, Academic Press, 2021.
- [82] Olivier Doré et al. “Cosmology with the SPHEREX All-Sky Spectral Survey”. In: (Dec. 2014). arXiv: [1412.4872](https://arxiv.org/abs/1412.4872) [[astro-ph.CO](#)].
- [83] Cyrille Doux et al. “Cosmological constraints from a joint analysis of cosmic microwave background and spectroscopic tracers of the large-scale structure”. In: 480.4 (2018), pp. 5386–5411. DOI: [10.1093/mnras/sty2160](https://doi.org/10.1093/mnras/sty2160). arXiv: [1706.04583](https://arxiv.org/abs/1706.04583) [[astro-ph.CO](#)].
- [84] J. Dunkley et al. “The Atacama Cosmology Telescope: likelihood for small-scale CMB data”. In: 2013.7, 025 (July 2013), p. 025. DOI: [10.1088/1475-7516/2013/07/025](https://doi.org/10.1088/1475-7516/2013/07/025). arXiv: [1301.0776](https://arxiv.org/abs/1301.0776) [[astro-ph.CO](#)].
- [85] Ruth Durrer. *The Cosmic Microwave Background*. 2nd ed. Cambridge University Press, 2020. DOI: [10.1017/9781316471524](https://doi.org/10.1017/9781316471524).
- [86] Daniel J. Eisenstein and Wayne Hu. “Power Spectra for Cold Dark Matter and Its Variants”. In: *The Astrophysical Journal* 511.1 (1999), 5–15. ISSN: 1538-4357. DOI: [10.1086/306640](https://doi.org/10.1086/306640). URL: <http://dx.doi.org/10.1086/306640>.
- [87] Hume A. Feldman, Nick Kaiser, and John A. Peacock. “Power-Spectrum Analysis of Three-dimensional Redshift Surveys”. In: 426 (May 1994), p. 23. DOI: [10.1086/174036](https://doi.org/10.1086/174036). arXiv: [astro-ph/9304022](https://arxiv.org/abs/astro-ph/9304022) [[astro-ph](#)].
- [88] J. R. Fergusson, D. M. Regan, and E. P. S. Shellard. “Rapid Separable Analysis of Higher Order Correlators in Large Scale Structure”. In: *Phys. Rev. D* 86 (2012), p. 063511. DOI: [10.1103/PhysRevD.86.063511](https://doi.org/10.1103/PhysRevD.86.063511). arXiv: [1008.1730](https://arxiv.org/abs/1008.1730) [[astro-ph.CO](#)].
- [89] Simone Ferraro and J. Colin Hill. “Bias to CMB lensing reconstruction from temperature anisotropies due to large-scale galaxy motions”. In: 97.2, 023512 (2018), p. 023512. DOI: [10.1103/PhysRevD.97.023512](https://doi.org/10.1103/PhysRevD.97.023512). arXiv: [1705.06751](https://arxiv.org/abs/1705.06751) [[astro-ph.CO](#)].
- [90] Simone Ferraro et al. “Inflation and Dark Energy from Spectroscopy at  $z > 2$ ”. In: (2019). arXiv: [1903.09208](https://arxiv.org/abs/1903.09208) [[astro-ph.CO](#)].
- [91] Simon Foreman et al. “Lensing reconstruction from line intensity maps: the impact of gravitational nonlinearity”. In: *JCAP* 1807.07 (2018), p. 046. DOI: [10.1088/1475-7516/2018/07/046](https://doi.org/10.1088/1475-7516/2018/07/046). arXiv: [1803.04975](https://arxiv.org/abs/1803.04975) [[astro-ph.CO](#)].
- [92] J. N. Fry and Enrique Gaztanaga. “Biasing and hierarchical statistics in large-scale structure”. In: *The Astrophysical Journal* 413 (1993), p. 447. ISSN: 1538-4357. DOI: [10.1086/173015](https://doi.org/10.1086/173015). URL: <http://dx.doi.org/10.1086/173015>.
- [93] J. E. Geach et al. “The Halo Mass of Optically Luminous Quasars at  $z \approx 1$ -2 Measured via Gravitational Deflection of the Cosmic Microwave Background”. In: 874.1, 85 (2019), p. 85. DOI: [10.3847/1538-4357/ab0894](https://doi.org/10.3847/1538-4357/ab0894). arXiv: [1902.06955](https://arxiv.org/abs/1902.06955) [[astro-ph.GA](#)].

- [94] Abhik Ghosh, Florent Mertens, and Leon V. E. Koopmans. “Deconvolving the wedge: maximum-likelihood power spectra via spherical-wave visibility modelling”. In: *Mon. Not. Roy. Astron. Soc.* 474.4 (2018), pp. 4552–4563. DOI: [10.1093/mnras/stx2959](https://doi.org/10.1093/mnras/stx2959). arXiv: [1709.06752](https://arxiv.org/abs/1709.06752) [[astro-ph.CO](#)].
- [95] T. Giannantonio et al. “CMB lensing tomography with the DES Science Verification galaxies”. In: *Monthly Notices of the Royal Astronomical Society* 456.3 (2016), pp. 3213–3244. DOI: [10.1093/mnras/stv2678](https://doi.org/10.1093/mnras/stv2678). eprint: [/oup/backfile/content\\_public/journal/mnras/456/3/10.1093\\_mnras\\_stv2678/1/stv2678.pdf](https://oup/backfile/content_public/journal/mnras/456/3/10.1093_mnras_stv2678/1/stv2678.pdf). URL: <http://dx.doi.org/10.1093/mnras/stv2678>.
- [96] T. Giannantonio et al. “CMB lensing tomography with the DES Science Verification galaxies”. In: 456.3 (2016), pp. 3213–3244. DOI: [10.1093/mnras/stv2678](https://doi.org/10.1093/mnras/stv2678). arXiv: [1507.05551](https://arxiv.org/abs/1507.05551) [[astro-ph.CO](#)].
- [97] Tommaso Giannantonio and Cristiano Porciani. “Structure formation from non-Gaussian initial conditions: multivariate biasing, statistics, and comparison with N-body simulations”. In: *Phys. Rev. D* 81 (2010), p. 063530. DOI: [10.1103/PhysRevD.81.063530](https://doi.org/10.1103/PhysRevD.81.063530). arXiv: [0911.0017](https://arxiv.org/abs/0911.0017) [[astro-ph.CO](#)].
- [98] Tommaso Giannantonio et al. “Improved Primordial Non-Gaussianity Constraints from Measurements of Galaxy Clustering and the Integrated Sachs-Wolfe Effect”. In: *Phys. Rev. D* 89.2 (2014), p. 023511. DOI: [10.1103/PhysRevD.89.023511](https://doi.org/10.1103/PhysRevD.89.023511). arXiv: [1303.1349](https://arxiv.org/abs/1303.1349) [[astro-ph.CO](#)].
- [99] Héctor Gil-Marín et al. “Dark matter and halo bispectrum in redshift space: theory and applications”. In: *JCAP* 12 (2014), p. 029. DOI: [10.1088/1475-7516/2014/12/029](https://doi.org/10.1088/1475-7516/2014/12/029). arXiv: [1407.1836](https://arxiv.org/abs/1407.1836) [[astro-ph.CO](#)].
- [100] Elena Giusarma et al. “Scale-dependent galaxy bias, CMB lensing-galaxy cross-correlation, and neutrino masses”. In: 98.12, 123526 (2018), p. 123526. DOI: [10.1103/PhysRevD.98.123526](https://doi.org/10.1103/PhysRevD.98.123526). arXiv: [1802.08694](https://arxiv.org/abs/1802.08694) [[astro-ph.CO](#)].
- [101] K. M. Górski et al. “HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere”. In: 622 (Apr. 2005), pp. 759–771. DOI: [10.1086/427976](https://doi.org/10.1086/427976). eprint: [arXiv:astro-ph/0409513](https://arxiv.org/abs/astro-ph/0409513).
- [102] Davide Gualdi et al. “Enhancing BOSS bispectrum cosmological constraints with maximal compression”. In: *Mon. Not. Roy. Astron. Soc.* 484.3 (2019), pp. 3713–3730. DOI: [10.1093/mnras/stz051](https://doi.org/10.1093/mnras/stz051). arXiv: [1806.02853](https://arxiv.org/abs/1806.02853) [[astro-ph.CO](#)].
- [103] Arushi Gupta et al. “Non-Gaussian information from weak lensing data via deep learning”. In: *Physical Review D* 97.10 (2018). ISSN: 2470-0029. DOI: [10.1103/PhysRevD.97.103515](https://doi.org/10.1103/PhysRevD.97.103515). URL: <http://dx.doi.org/10.1103/PhysRevD.97.103515>.
- [104] Alan H. Guth. “THE INFLATIONARY UNIVERSE: A POSSIBLE SOLUTION TO THE HORIZON AND FLATNESS PROBLEMS”. In: *Phys. Rev. D* 23 (1981), pp. 347–356.
- [105] Nico Hamaus, Uros Seljak, and Vincent Desjacques. “Optimal Constraints on Local Primordial Non-Gaussianity from the Two-Point Statistics of Large-Scale Structure”. In: *Phys. Rev. D* 84 (2011), p. 083509. DOI: [10.1103/PhysRevD.84.083509](https://doi.org/10.1103/PhysRevD.84.083509). arXiv: [1104.2321](https://arxiv.org/abs/1104.2321) [[astro-ph.CO](#)].
- [106] Nico Hamaus et al. “Minimizing the stochasticity of halos in large-scale structure surveys”. In: 82.4, 043515 (Aug. 2010), p. 043515. DOI: [10.1103/PhysRevD.82.043515](https://doi.org/10.1103/PhysRevD.82.043515). arXiv: [1004.5377](https://arxiv.org/abs/1004.5377) [[astro-ph.CO](#)].

- [107] Jiashu Han et al. “Probing gravitational lensing of the CMB with SDSS-IV quasars”. In: 485.2 (2019), pp. 1720–1726. DOI: [10.1093/mnras/stz528](https://doi.org/10.1093/mnras/stz528). arXiv: [1809.04196](https://arxiv.org/abs/1809.04196) [astro-ph.CO].
- [108] Nick Hand et al. “First measurement of the cross-correlation of CMB lensing and galaxy lensing”. In: 91.6, 062001 (Mar. 2015), p. 062001. DOI: [10.1103/PhysRevD.91.062001](https://doi.org/10.1103/PhysRevD.91.062001). arXiv: [1311.6200](https://arxiv.org/abs/1311.6200) [astro-ph.CO].
- [109] Duncan Hanson et al. “CMB temperature lensing power reconstruction”. In: 83.4, 043005 (2011), p. 043005. DOI: [10.1103/PhysRevD.83.043005](https://doi.org/10.1103/PhysRevD.83.043005). arXiv: [1008.4403](https://arxiv.org/abs/1008.4403) [astro-ph.CO].
- [110] Duncan Hanson et al. “CMB temperature lensing power reconstruction”. In: *Phys. Rev. D* 83 (2011), p. 043005. DOI: [10.1103/PhysRevD.83.043005](https://doi.org/10.1103/PhysRevD.83.043005). arXiv: [1008.4403](https://arxiv.org/abs/1008.4403) [astro-ph.CO].
- [111] J. Hartlap, P. Simon, and P. Schneider. “Why your model parameter confidences might be too optimistic. Unbiased estimation of the inverse covariance matrix”. In: *A&A* 464.1 (2007), pp. 399–404.
- [112] J. Colin Hill. “Foreground biases on primordial non-Gaussianity measurements from the CMB temperature bispectrum: Implications for Planck and beyond”. In: *Physical Review D* 98.8 (2018). ISSN: 2470-0029. DOI: [10.1103/physrevd.98.083542](https://doi.org/10.1103/physrevd.98.083542). URL: <http://dx.doi.org/10.1103/PhysRevD.98.083542>.
- [113] M. Hilton et al. “The Atacama Cosmology Telescope: A Catalog of >4000 Sunyaev–Zel’dovich Galaxy Clusters”. In: *The Astrophysical Journal Supplement Series* 253.1 (2021), p. 3. ISSN: 1538-4365. DOI: [10.3847/1538-4365/abd023](https://doi.org/10.3847/1538-4365/abd023). URL: <http://dx.doi.org/10.3847/1538-4365/abd023>.
- [114] Christopher M. Hirata and Uro š Seljak. “Reconstruction of lensing from the cosmic microwave background polarization”. In: *Phys. Rev. D* 68 (8 2003), p. 083002. DOI: [10.1103/PhysRevD.68.083002](https://doi.org/10.1103/PhysRevD.68.083002). URL: <https://link.aps.org/doi/10.1103/PhysRevD.68.083002>.
- [115] Christopher M. Hirata et al. “Correlation of CMB with large-scale structure. II. Weak lensing”. In: *Physical Review D* 78.4 (2008). ISSN: 1550-2368. DOI: [10.1103/physrevd.78.043520](https://doi.org/10.1103/physrevd.78.043520). URL: <http://dx.doi.org/10.1103/PhysRevD.78.043520>.
- [116] Christopher M. Hirata et al. “Correlation of CMB with large-scale structure. II. Weak lensing”. In: 78.4, 043520 (2008), p. 043520. DOI: [10.1103/PhysRevD.78.043520](https://doi.org/10.1103/PhysRevD.78.043520). arXiv: [0801.0644](https://arxiv.org/abs/0801.0644) [astro-ph].
- [117] Christopher M. Hirata et al. “Cross-correlation of CMB with large-scale structure: Weak gravitational lensing”. In: *Physical Review D* 70.10 (2004). ISSN: 1550-2368. DOI: [10.1103/physrevd.70.103501](https://doi.org/10.1103/physrevd.70.103501). URL: <http://dx.doi.org/10.1103/PhysRevD.70.103501>.
- [118] Christopher M. Hirata et al. “Cross-correlation of CMB with large-scale structure: Weak gravitational lensing”. In: 70.10, 103501 (2004), p. 103501. DOI: [10.1103/PhysRevD.70.103501](https://doi.org/10.1103/PhysRevD.70.103501). arXiv: [astro-ph/0406004](https://arxiv.org/abs/astro-ph/0406004) [astro-ph].
- [119] Shirley Ho et al. “Sloan Digital Sky Survey III photometric quasar clustering: probing the initial conditions of the Universe”. In: *JCAP* 05 (2015), p. 040. DOI: [10.1088/1475-7516/2015/05/040](https://doi.org/10.1088/1475-7516/2015/05/040). arXiv: [1311.2597](https://arxiv.org/abs/1311.2597) [astro-ph.CO].

- [120] G. P. Holder et al. “A Cosmic Microwave Background Lensing Mass Map and Its Correlation with the Cosmic Infrared Background”. In: 771.1, L16 (July 2013), p. L16. DOI: [10.1088/2041-8205/771/1/L16](https://doi.org/10.1088/2041-8205/771/1/L16). arXiv: [1303.5048](https://arxiv.org/abs/1303.5048) [[astro-ph.CO](#)].
- [121] Wayne Hu. “Concepts in CMB anisotropy formation”. In: *Lecture Notes in Physics* (), 207–239. DOI: [10.1007/bfb0102588](https://doi.org/10.1007/bfb0102588). URL: <http://dx.doi.org/10.1007/BFb0102588>.
- [122] Wayne Hu. “Concepts in CMB anisotropy formation”. In: *Lecture Notes in Physics* (), 207–239. DOI: [10.1007/bfb0102588](https://doi.org/10.1007/bfb0102588). URL: <http://dx.doi.org/10.1007/BFb0102588>.
- [123] Wayne Hu. “Weak lensing of the CMB: A harmonic approach”. In: *Physical Review D - Particles, Fields, Gravitation and Cosmology* 62.4 (2000), p. 17. ISSN: 15502368. DOI: [10.1103/PhysRevD.62.043007](https://doi.org/10.1103/PhysRevD.62.043007). arXiv: [0001303](https://arxiv.org/abs/0001303) [[astro-ph](#)].
- [124] Wayne Hu, Simon DeDeo, and Chris Vale. “Cluster mass estimators from CMB temperature and polarization lensing”. In: *New Journal of Physics* 9.12 (2007), p. 441. URL: <http://stacks.iop.org/1367-2630/9/i=12/a=441>.
- [125] Wayne Hu, Simon DeDeo, and Chris Vale. “Cluster mass estimators from CMB temperature and polarization lensing”. In: *New Journal of Physics* 9.12 (2007), p. 441. DOI: [10.1088/1367-2630/9/12/441](https://doi.org/10.1088/1367-2630/9/12/441). arXiv: [astro-ph/0701276](https://arxiv.org/abs/astro-ph/0701276) [[astro-ph](#)].
- [126] Wayne Hu and Scott Dodelson. “Cosmic Microwave Background Anisotropies”. In: *Annual Review of Astronomy and Astrophysics* 40.1 (2002), 171–216. ISSN: 1545-4282. DOI: [10.1146/annurev.astro.40.060401.093926](https://doi.org/10.1146/annurev.astro.40.060401.093926). URL: <http://dx.doi.org/10.1146/annurev.astro.40.060401.093926>.
- [127] Wayne Hu, Daniel E. Holz, and Chris Vale. “CMB cluster lensing: Cosmography with the longest lever arm”. In: 76.12, 127301 (2007), p. 127301. DOI: [10.1103/PhysRevD.76.127301](https://doi.org/10.1103/PhysRevD.76.127301). arXiv: [0708.4391](https://arxiv.org/abs/0708.4391) [[astro-ph](#)].
- [128] Wayne Hu and Takemi Okamoto. “Mass reconstruction with cmb polarization”. In: *Astrophys. J.* 574 (2002), pp. 566–574. DOI: [10.1086/341110](https://doi.org/10.1086/341110). arXiv: [astro-ph/0111606](https://arxiv.org/abs/astro-ph/0111606).
- [129] Wayne Hu and Takemi Okamoto. “Mass Reconstruction with Cosmic Microwave Background Polarization”. In: *The Astrophysical Journal* 574.2 (2002), 566–574. ISSN: 1538-4357. DOI: [10.1086/341110](https://doi.org/10.1086/341110). URL: <http://dx.doi.org/10.1086/341110>.
- [130] Wayne Hu and Takemi Okamoto. “Mass Reconstruction with Cosmic Microwave Background Polarization”. In: 574.2 (2002), pp. 566–574. DOI: [10.1086/341110](https://doi.org/10.1086/341110). arXiv: [astro-ph/0111606](https://arxiv.org/abs/astro-ph/0111606) [[astro-ph](#)].
- [131] Wayne Hu and Martin White. “A CMB polarization primer”. In: *New Astronomy* 2.4 (1997), 323–344. ISSN: 1384-1076. DOI: [10.1016/s1384-1076\(97\)00022-5](https://doi.org/10.1016/s1384-1076(97)00022-5). URL: [http://dx.doi.org/10.1016/S1384-1076\(97\)00022-5](http://dx.doi.org/10.1016/S1384-1076(97)00022-5).
- [132] J. D. Hunter. “Matplotlib: A 2D graphics environment”. In: *Computing in Science & Engineering* 9.3 (2007), pp. 90–95. DOI: [10.1109/MCSE.2007.55](https://doi.org/10.1109/MCSE.2007.55).
- [133] G. Hurier, P. Singh, and C. Hernández-Monteagudo. “First measurement of the cross-correlation between CMB weak lensing and X-ray emission”. In: *arXiv e-prints*, arXiv:1711.10774 (2017), arXiv:1711.10774. arXiv: [1711.10774](https://arxiv.org/abs/1711.10774) [[astro-ph.CO](#)].

- [134] Mikhail M. Ivanov, Marko Simonovic, and Matias Zaldarriaga. “Cosmological Parameters from the BOSS Galaxy Power Spectrum”. In: (Sept. 2019). arXiv: 1909.05277 [astro-ph.CO].
- [135] A. Jenkins et al. “The mass function of dark matter haloes”. In: 321.2 (Feb. 2001), pp. 372–384. DOI: 10.1046/j.1365-8711.2001.04029.x. arXiv: astro-ph/0005260 [astro-ph].
- [136] D. Jeong. [http://www.personal.psu.edu/duj13/dissertation/djeong\\_diss.pdf](http://www.personal.psu.edu/duj13/dissertation/djeong_diss.pdf). 2010.
- [137] Donghui Jeong and Marc Kamionkowski. “Clustering Fossils from the Early Universe”. In: *Phys. Rev. Lett.* 108 (2012), p. 251301. DOI: 10.1103/PhysRevLett.108.251301. arXiv: 1203.0302 [astro-ph.CO].
- [138] N. Kaiser. “Clustering in real space and in redshift space”. In: *Mon. Not. Roy. Astron. Soc.* 227 (1987), pp. 1–27.
- [139] B. Kalus et al. “A map-based method for eliminating systematic modes from galaxy clustering power spectra with application to BOSS”. In: *Mon. Not. Roy. Astron. Soc.* 482.1 (2019), pp. 453–470. DOI: 10.1093/mnras/sty2655. arXiv: 1806.02789 [astro-ph.CO].
- [140] Manoj Kaplinghat, Lloyd Knox, and Yong-Seon Song. “Determining Neutrino Mass from the Cosmic Microwave Background Alone”. In: *Physical Review Letters* 91.24 (2003). ISSN: 1079-7114. DOI: 10.1103/physrevlett.91.241301. URL: <http://dx.doi.org/10.1103/PhysRevLett.91.241301>.
- [141] Naim Göksel Karaçaylı and Nikhil Padmanabhan. “Anatomy of Cosmic Tidal Reconstruction”. In: *Mon. Not. Roy. Astron. Soc.* 486.3 (2019), pp. 3864–3873. DOI: 10.1093/mnras/stz964. arXiv: 1904.01387 [astro-ph.CO].
- [142] Issha Kayo and Masahiro Takada. “Cosmological parameters from weak lensing power spectrum and bispectrum tomography: including the non-Gaussian errors”. In: (June 2013). arXiv: 1306.4684 [astro-ph.CO].
- [143] Jaiseung Kim, Aditya Rotti, and Eiichiro Komatsu. “Removing the ISW-lensing bias from the local-form primordial non-Gaussianity estimation”. In: *Journal of Cosmology and Astroparticle Physics* 2013.04 (2013), 021–021. ISSN: 1475-7516. DOI: 10.1088/1475-7516/2013/04/021. URL: <http://dx.doi.org/10.1088/1475-7516/2013/04/021>.
- [144] Edward W Kolb and Michael Stanley Turner. *The early universe*. Frontiers in physics. Boulder, CO: Westview Press, 1990. DOI: 10.1201/9780429492860. URL: <https://cds.cern.ch/record/206230>.
- [145] Eiichiro Komatsu. “Critical Tests of Inflation as a Mechanism for Generating Observed Cosmological Fluctuations in the Universe”. In: URL: [https://wwwmpa.mpa-garching.mpg.de/~komatsu/cmb/komatsu\\_jgrg21\\_proceedings.pdf](https://wwwmpa.mpa-garching.mpg.de/~komatsu/cmb/komatsu_jgrg21_proceedings.pdf).
- [146] Eiichiro Komatsu. “Some Basics of the Expansion of the Universe, Cosmic Microwave Background, and Large-scale Structure of the Universe”. In: (). URL: [https://wwwmpa.mpa-garching.mpg.de/~komatsu/cmb/lecture\\_cosmo\\_iucaa\\_2011.pdf](https://wwwmpa.mpa-garching.mpg.de/~komatsu/cmb/lecture_cosmo_iucaa_2011.pdf).

- [147] D. Larson et al. “SEVEN-YEAR WILKINSON MICROWAVE ANISOTROPY PROBE ( WMAP ) OBSERVATIONS: POWER SPECTRA AND WMAP-DERIVED PARAMETERS”. In: *The Astrophysical Journal Supplement Series* 192.2 (2011), p. 16. ISSN: 1538-4365. DOI: [10.1088/0067-0049/192/2/16](https://doi.org/10.1088/0067-0049/192/2/16). URL: <http://dx.doi.org/10.1088/0067-0049/192/2/16>.
- [148] Titouan Lazeyras et al. “Precision measurement of the local bias of dark matter halos”. In: *JCAP* 1602.02 (2016), p. 018. DOI: [10.1088/1475-7516/2016/02/018](https://doi.org/10.1088/1475-7516/2016/02/018). arXiv: [1511.01096](https://arxiv.org/abs/1511.01096) [[astro-ph.CO](https://arxiv.org/abs/1511.01096)].
- [149] Florent Leclercq, Alice Pisani, and Benjamin D. Wandelt. “Cosmology: from theory to data, from data to theory”. In: *Proc. Int. Sch. Phys. Fermi* 186 (2014). Ed. by A. Cooray et al., pp. 189–233. DOI: [10.3254/978-1-61499-476-3-189](https://doi.org/10.3254/978-1-61499-476-3-189). arXiv: [1403.1260](https://arxiv.org/abs/1403.1260) [[astro-ph.CO](https://arxiv.org/abs/1403.1260)].
- [150] Boris Leistedt, Hiranya V. Peiris, and Nina Roth. “Constraints on Primordial Non-Gaussianity from 800 000 Photometric Quasars”. In: *Phys. Rev. Lett.* 113.22 (2014), p. 221301. DOI: [10.1103/PhysRevLett.113.221301](https://doi.org/10.1103/PhysRevLett.113.221301). arXiv: [1405.4315](https://arxiv.org/abs/1405.4315) [[astro-ph.CO](https://arxiv.org/abs/1405.4315)].
- [151] J. Lesgourgues and L. Verde. “Neutrinos in Cosmology”. In: (). URL: <https://pdg.lbl.gov/2020/reviews/rpp2020-rev-neutrinos-in-cosmology.pdf>.
- [152] Julien Lesgourgues et al. “Probing neutrino masses with CMB lensing extraction”. In: 73.4, 045021 (2006), p. 045021. DOI: [10.1103/PhysRevD.73.045021](https://doi.org/10.1103/PhysRevD.73.045021). arXiv: [astro-ph/0511735](https://arxiv.org/abs/astro-ph/0511735) [[astro-ph](https://arxiv.org/abs/astro-ph/0511735)].
- [153] Antony Lewis and Anthony Challinor. “Weak gravitational lensing of the CMB”. In: 429.1 (2006), pp. 1–65. DOI: [10.1016/j.physrep.2006.03.002](https://doi.org/10.1016/j.physrep.2006.03.002). arXiv: [astro-ph/0601594](https://arxiv.org/abs/astro-ph/0601594) [[astro-ph](https://arxiv.org/abs/astro-ph/0601594)].
- [154] Antony Lewis, Anthony Challinor, and Anthony Lasenby. “Efficient Computation of Cosmic Microwave Background Anisotropies in Closed Friedmann-Robertson-Walker Models”. In: 538.2 (2000), pp. 473–476. DOI: [10.1086/309179](https://doi.org/10.1086/309179). arXiv: [astro-ph/9911177](https://arxiv.org/abs/astro-ph/9911177) [[astro-ph](https://arxiv.org/abs/astro-ph/9911177)].
- [155] Dongzi Li, Hong-Ming Zhu, and Ue-Li Pen. “Cross-correlation of the kinematic Sunyaev-Zel’dovich effect and 21 cm intensity mapping with tidal reconstruction”. In: *Phys. Rev. D* 100.2 (2019), p. 023517. DOI: [10.1103/PhysRevD.100.023517](https://doi.org/10.1103/PhysRevD.100.023517). arXiv: [1811.05012](https://arxiv.org/abs/1811.05012) [[astro-ph.CO](https://arxiv.org/abs/1811.05012)].
- [156] Peikai Li, Rupert A.C. Croft, and Scott Dodelson. “Large Scale Structure Reconstruction with Short-Wavelength Modes: Halo Bias and Light Cone Formalism”. In: (July 2020). arXiv: [2007.00226](https://arxiv.org/abs/2007.00226) [[astro-ph.CO](https://arxiv.org/abs/2007.00226)].
- [157] Peikai Li, Scott Dodelson, and Rupert A.C. Croft. “Large Scale Structure Reconstruction with Short-Wavelength Modes”. In: *Phys. Rev. D* 101.8 (2020), p. 083510. DOI: [10.1103/PhysRevD.101.083510](https://doi.org/10.1103/PhysRevD.101.083510). arXiv: [2001.02780](https://arxiv.org/abs/2001.02780) [[astro-ph.CO](https://arxiv.org/abs/2001.02780)].
- [158] D. N. Limber. “The Analysis of Counts of the Extragalactic Nebulae in Terms of a Fluctuating Density Field.” In: 117 (1953), p. 134. DOI: [10.1086/145672](https://doi.org/10.1086/145672).
- [159] A. Linde. “A NEW INFLATIONARY UNIVERSE SCENARIO: A POSSIBLE SOLUTION OF THE HORIZON, FLATNESS, HOMOGENEITY, ISOTROPY AND PRIMORDIAL MONOPOLE PROBLEMS”. In: *Phys. Lett.* 108B (1982), pp. 389–393.
- [160] A. Linde. “Scalar Field Fluctuations in Expanding Universe and the New Inflationary Universe Scenario”. In: *Phys. Lett.* 116B (1982), p. 335.

- [161] Adrian Liu, Aaron R. Parsons, and Cathryn M. Trott. “Epoch of reionization window. I. Mathematical formalism”. In: *Phys. Rev. D* 90.2 (2014), p. 023018. DOI: [10.1103/PhysRevD.90.023018](https://doi.org/10.1103/PhysRevD.90.023018). arXiv: [1404.2596](https://arxiv.org/abs/1404.2596) [[astro-ph.CO](#)].
- [162] Adrian Liu, Aaron R. Parsons, and Cathryn M. Trott. “Epoch of reionization window. II. Statistical methods for foreground wedge reduction”. In: *Phys. Rev. D* 90.2 (2014), p. 023019. DOI: [10.1103/PhysRevD.90.023019](https://doi.org/10.1103/PhysRevD.90.023019). arXiv: [1404.4372](https://arxiv.org/abs/1404.4372) [[astro-ph.CO](#)].
- [163] Jia Liu and J. Colin Hill. “Cross-correlation of Planck CMB lensing and CFHTLenS galaxy weak lensing maps”. In: 92.6, 063517 (2015), p. 063517. DOI: [10.1103/PhysRevD.92.063517](https://doi.org/10.1103/PhysRevD.92.063517). arXiv: [1504.05598](https://arxiv.org/abs/1504.05598) [[astro-ph.CO](#)].
- [164] Jia Liu, Alvaro Ortiz-Vazquez, and J. Colin Hill. “Constraining multiplicative bias in CFHTLenS weak lensing shear data”. In: 93.10, 103508 (2016), p. 103508. DOI: [10.1103/PhysRevD.93.103508](https://doi.org/10.1103/PhysRevD.93.103508). arXiv: [1601.05720](https://arxiv.org/abs/1601.05720) [[astro-ph.CO](#)].
- [165] R. Henry Liu and Patrick C. Breysse. “Coupling parsec and gigaparsec scales: primordial non-Gaussianity with multi-tracer intensity mapping”. In: (2020). arXiv: [2002.10483](https://arxiv.org/abs/2002.10483) [[astro-ph.CO](#)].
- [166] Thibaut Louis et al. “Lensing simulation and power spectrum estimation for high-resolution CMB polarization maps”. In: 435.3 (2013), pp. 2040–2047. DOI: [10.1093/mnras/stt1421](https://doi.org/10.1093/mnras/stt1421). arXiv: [1306.6692](https://arxiv.org/abs/1306.6692) [[astro-ph.CO](#)].
- [167] David H. Lyth and Andrew R. Liddle. *The Primordial Density Perturbation: Cosmology, Inflation and the Origin of Structure*. Cambridge University Press, 2009. DOI: [10.1017/CBO9780511819209](https://doi.org/10.1017/CBO9780511819209).
- [168] Mathew Madhavacheril et al. “Evidence of Lensing of the Cosmic Microwave Background by Dark Matter Halos”. In: 114.15, 151302 (2015), p. 151302. DOI: [10.1103/PhysRevLett.114.151302](https://doi.org/10.1103/PhysRevLett.114.151302). arXiv: [1411.7999](https://arxiv.org/abs/1411.7999) [[astro-ph.CO](#)].
- [169] Mathew S. Madhavacheril and J. Colin Hill. “Mitigating foreground biases in CMB lensing reconstruction using cleaned gradients”. In: 98.2, 023534 (2018), p. 023534. DOI: [10.1103/PhysRevD.98.023534](https://doi.org/10.1103/PhysRevD.98.023534). arXiv: [1802.08230](https://arxiv.org/abs/1802.08230) [[astro-ph.CO](#)].
- [170] Mathew S. Madhavacheril et al. “Atacama Cosmology Telescope: Component-separated maps of CMB temperature and the thermal Sunyaev-Zel’dovich effect”. In: 102.2, 023534 (July 2020), p. 023534. DOI: [10.1103/PhysRevD.102.023534](https://doi.org/10.1103/PhysRevD.102.023534). arXiv: [1911.05717](https://arxiv.org/abs/1911.05717) [[astro-ph.CO](#)].
- [171] Mathew S. Madhavacheril et al. “Atacama Cosmology Telescope: Component-separated maps of CMB temperature and the thermal Sunyaev-Zel’dovich effect”. In: *Physical Review D* 102.2 (2020). ISSN: 2470-0029. DOI: [10.1103/PhysRevD.102.023534](https://doi.org/10.1103/PhysRevD.102.023534). URL: <http://dx.doi.org/10.1103/PhysRevD.102.023534>.
- [172] Juan Maldacena. “Non-gaussian features of primordial fluctuations in single field inflationary models”. In: *Journal of High Energy Physics* 2003.05 (2003), 013–013. ISSN: 1029-8479. DOI: [10.1088/1126-6708/2003/05/013](https://doi.org/10.1088/1126-6708/2003/05/013). URL: <http://dx.doi.org/10.1088/1126-6708/2003/05/013>.
- [173] Kiyoshi Wesley Masui and Ue-Li Pen. “Primordial gravity wave fossils and their use in testing inflation”. In: *Phys. Rev. Lett.* 105 (2010), p. 161302. DOI: [10.1103/PhysRevLett.105.161302](https://doi.org/10.1103/PhysRevLett.105.161302). arXiv: [1006.4181](https://arxiv.org/abs/1006.4181) [[astro-ph.CO](#)].

- [174] Kiyoshi Wesley Masui, Ue-Li Pen, and Neil Turok. “Two- and Three-Dimensional Probes of Parity in Primordial Gravity Waves”. In: *Phys. Rev. Lett.* 118.22 (2017), p. 221301. DOI: [10.1103/PhysRevLett.118.221301](https://doi.org/10.1103/PhysRevLett.118.221301). arXiv: [1702.06552](https://arxiv.org/abs/1702.06552) [[astro-ph.CO](#)].
- [175] Sabino Matarrese. *Les Houches Cosmic Microwave Background lecture notes*. 2006.
- [176] Sabino Matarrese and Licia Verde. “The effect of primordial non-Gaussianity on halo bias”. In: *Astrophys. J. Lett.* 677 (2008), pp. L77–L80. DOI: [10.1086/587840](https://doi.org/10.1086/587840). arXiv: [0801.4826](https://arxiv.org/abs/0801.4826) [[astro-ph](#)].
- [177] Fiona McCarthy and Mathew S. Madhavacheril. “Improving models of the cosmic infrared background using CMB lensing mass maps”. In: *Physical Review D* 103.10 (2021). ISSN: 2470-0029. DOI: [10.1103/physrevd.103.103515](https://doi.org/10.1103/physrevd.103.103515). URL: <http://dx.doi.org/10.1103/PhysRevD.103.103515>.
- [178] Patrick McDonald. “Clustering of dark matter tracers: renormalizing the bias parameters”. In: (2006). arXiv: [astro-ph/0609413](https://arxiv.org/abs/astro-ph/0609413) [[astro-ph](#)].
- [179] Patrick McDonald and Uros Seljak. “How to measure redshift-space distortions without sample variance”. In: *JCAP* 0910 (2009), p. 007. DOI: [10.1088/1475-7516/2009/10/007](https://doi.org/10.1088/1475-7516/2009/10/007). arXiv: [0810.0323](https://arxiv.org/abs/0810.0323) [[astro-ph](#)].
- [180] Michael McKerns, Patrick Hung, and Michael Aivazis. *mystic: highly-constrained non-convex optimization and UQ*. URL: <https://uqfoundation.github.io/project/mystic>.
- [181] Michael M. McKerns et al. *Building a Framework for Predictive Science*. 2012. arXiv: [1202.1056](https://arxiv.org/abs/1202.1056) [[cs.MS](#)].
- [182] A. J. Mead et al. “An accurate halo model for fitting non-linear cosmological power spectra and baryonic feedback models”. In: 454.2 (2015), pp. 1958–1975. DOI: [10.1093/mnras/stv2036](https://doi.org/10.1093/mnras/stv2036). arXiv: [1505.07833](https://arxiv.org/abs/1505.07833) [[astro-ph.CO](#)].
- [183] P. Daniel Meerburg et al. “Primordial Non-Gaussianity”. In: (Mar. 2019). arXiv: [1903.04409](https://arxiv.org/abs/1903.04409) [[astro-ph.CO](#)].
- [184] Jean-Baptiste Melin and James G. Bartlett. “Measuring cluster masses with CMB lensing: a statistical approach”. In: 578, A21 (2015), A21. DOI: [10.1051/0004-6361/201424720](https://doi.org/10.1051/0004-6361/201424720). arXiv: [1408.5633](https://arxiv.org/abs/1408.5633) [[astro-ph.CO](#)].
- [185] M. Millea et al. “MODELING EXTRAGALACTIC FOREGROUNDS AND SECONDARIES FOR UNBIASED ESTIMATION OF COSMOLOGICAL PARAMETERS FROM PRIMARY COSMIC MICROWAVE BACKGROUND ANISOTROPY”. In: *The Astrophysical Journal* 746.1 (2012), p. 4. DOI: [10.1088/0004-637x/746/1/4](https://doi.org/10.1088/0004-637x/746/1/4). URL: <https://doi.org/10.1088/0004-637x/746/1/4>.
- [186] Hironao Miyatake et al. “Measurement of a Cosmographic Distance Ratio with Galaxy and Cosmic Microwave Background Lensing”. In: 118.16, 161301 (2017), p. 161301. DOI: [10.1103/PhysRevLett.118.161301](https://doi.org/10.1103/PhysRevLett.118.161301). arXiv: [1605.05337](https://arxiv.org/abs/1605.05337) [[astro-ph.CO](#)].
- [187] Chirag Modi, Martin White, and Zvonimir Vlah. “Modeling CMB lensing cross correlations with CLEFT”. In: *Journal of Cosmology and Astroparticle Physics* 2017.08 (2017), 009–009. ISSN: 1475-7516. DOI: [10.1088/1475-7516/2017/08/009](https://doi.org/10.1088/1475-7516/2017/08/009). URL: <http://dx.doi.org/10.1088/1475-7516/2017/08/009>.

- [188] Chirag Modi et al. “Reconstructing large-scale structure with neutral hydrogen surveys”. In: *JCAP* 1911.11 (2019), p. 023. DOI: [10.1088/1475-7516/2019/11/023](https://doi.org/10.1088/1475-7516/2019/11/023). arXiv: [1907.02330](https://arxiv.org/abs/1907.02330) [[astro-ph.CO](#)].
- [189] Azadeh Moradinezhad Dizgah et al. “Capturing Non-Gaussianity of the Large-Scale Structure with Weighted Skew-Spectra”. In: (2019). arXiv: [1911.05763](https://arxiv.org/abs/1911.05763) [[astro-ph.CO](#)].
- [190] V.F. Mukhanov, H.A. Feldman, and R.H. Brandenberger. “Theory of cosmological perturbations”. In: *Physics Reports* 215.5 (1992), pp. 203–333. ISSN: 0370-1573. DOI: [https://doi.org/10.1016/0370-1573\(92\)90044-Z](https://doi.org/10.1016/0370-1573(92)90044-Z). URL: <https://www.sciencedirect.com/science/article/pii/037015739290044Z>.
- [191] Viatcheslav Mukhanov. *Physical Foundations of Cosmology*. Cambridge University Press, 2005. DOI: [10.1017/CBO9780511790553](https://doi.org/10.1017/CBO9780511790553).
- [192] Moritz Münchmeyer et al. “Constraining local non-Gaussianities with kinetic Sunyaev-Zel’dovich tomography”. In: *Phys. Rev. D* 100.8 (2019), p. 083508. DOI: [10.1103/PhysRevD.100.083508](https://doi.org/10.1103/PhysRevD.100.083508). arXiv: [1810.13424](https://arxiv.org/abs/1810.13424) [[astro-ph.CO](#)].
- [193] Sigurd Naess et al. “The Atacama Cosmology Telescope: arcminute-resolution maps of 18 000 square degrees of the microwave sky from ACT 2008–2018 data combined with Planck”. In: *Journal of Cosmology and Astroparticle Physics* 2020.12 (2020), 046–046. ISSN: 1475-7516. DOI: [10.1088/1475-7516/2020/12/046](https://doi.org/10.1088/1475-7516/2020/12/046). URL: <http://dx.doi.org/10.1088/1475-7516/2020/12/046>.
- [194] Toshiya Namikawa. “Cosmology from weak lensing of CMB”. In: *Progress of Theoretical and Experimental Physics* 2014.6 (June 2014). 06B108. ISSN: 2050-3911. DOI: [10.1093/ptep/ptu044](https://doi.org/10.1093/ptep/ptu044). eprint: <https://academic.oup.com/ptep/article-pdf/2014/6/06B108/4443636/ptu044.pdf>. URL: <https://doi.org/10.1093/ptep/ptu044>.
- [195] Toshiya Namikawa, Duncan Hanson, and Ryuichi Takahashi. “Bias-hardened CMB lensing”. In: 431.1 (2013), pp. 609–620. DOI: [10.1093/mnras/stt195](https://doi.org/10.1093/mnras/stt195). arXiv: [1209.0091](https://arxiv.org/abs/1209.0091) [[astro-ph.CO](#)].
- [196] Toshiya Namikawa, Duncan Hanson, and Ryuichi Takahashi. “Bias-hardened CMB lensing”. In: *Monthly Notices of the Royal Astronomical Society* 431.1 (2013), 609–620. ISSN: 1365-2966. DOI: [10.1093/mnras/stt195](https://doi.org/10.1093/mnras/stt195). URL: <http://dx.doi.org/10.1093/mnras/stt195>.
- [197] Toshiya Namikawa, Duncan Hanson, and Ryuichi Takahashi. “Bias-Hardened CMB Lensing”. In: *Mon. Not. Roy. Astron. Soc.* 431 (2013), pp. 609–620. DOI: [10.1093/mnras/stt195](https://doi.org/10.1093/mnras/stt195). arXiv: [1209.0091](https://arxiv.org/abs/1209.0091) [[astro-ph.CO](#)].
- [198] Y. Omori and G. Holder. “Cross-Correlation of CFHTLenS Galaxy Number Density and Planck CMB Lensing”. In: *arXiv e-prints*, arXiv:1502.03405 (Feb. 2015), arXiv:1502.03405. arXiv: [1502.03405](https://arxiv.org/abs/1502.03405) [[astro-ph.CO](#)].
- [199] Y. Omori et al. “A 2500 deg<sup>2</sup> CMB Lensing Map from Combined South Pole Telescope and Planck Data”. In: 849.2, 124 (Nov. 2017), p. 124. DOI: [10.3847/1538-4357/aa8d1d](https://doi.org/10.3847/1538-4357/aa8d1d). arXiv: [1705.00743](https://arxiv.org/abs/1705.00743) [[astro-ph.CO](#)].
- [200] Y. Omori et al. “Dark Energy Survey Year 1 Results: Tomographic cross-correlations between Dark Energy Survey galaxies and CMB lensing from South Pole Telescope +Planck”. In: 100.4, 043501 (2019), p. 043501. DOI: [10.1103/PhysRevD.100.043501](https://doi.org/10.1103/PhysRevD.100.043501). arXiv: [1810.02342](https://arxiv.org/abs/1810.02342) [[astro-ph.CO](#)].

- [201] Stephen J. Osborne, Duncan Hanson, and Olivier Doré. “Extragalactic Foreground Contamination in Temperature-based CMB Lens Reconstruction”. In: *JCAP* 1403 (2014), p. 024. DOI: [10.1088/1475-7516/2014/03/024](https://doi.org/10.1088/1475-7516/2014/03/024). arXiv: [1310.7547](https://arxiv.org/abs/1310.7547) [[astro-ph.CO](#)].
- [202] Stephen J. Osborne, Duncan Hanson, and Olivier Doré. “Extragalactic foreground contamination in temperature-based CMB lens reconstruction”. In: 2014.3, 024 (2014), p. 024. DOI: [10.1088/1475-7516/2014/03/024](https://doi.org/10.1088/1475-7516/2014/03/024). arXiv: [1310.7547](https://arxiv.org/abs/1310.7547) [[astro-ph.CO](#)].
- [203] Z. Pan, L. Knox, and M. White. “Dependence of the cosmic microwave background lensing power spectrum on the matter density”. In: *Monthly Notices of the Royal Astronomical Society* 445.3 (2014), 2941–2945. ISSN: 0035-8711. DOI: [10.1093/mnras/stu1971](https://doi.org/10.1093/mnras/stu1971). URL: <http://dx.doi.org/10.1093/mnras/stu1971>.
- [204] Aaron R. Parsons et al. “A Per-Baseline, Delay-Spectrum Technique for Accessing the 21cm Cosmic Reionization Signature”. In: *Astrophys. J.* 756 (2012), p. 165. DOI: [10.1088/0004-637X/756/2/165](https://doi.org/10.1088/0004-637X/756/2/165). arXiv: [1204.4749](https://arxiv.org/abs/1204.4749) [[astro-ph.IM](#)].
- [205] J. A. Peacock and M. Bilicki. “Wide-area tomography of CMB lensing and the growth of cosmological density fluctuations”. In: 481.1 (2018), pp. 1133–1148. DOI: [10.1093/mnras/sty2314](https://doi.org/10.1093/mnras/sty2314). arXiv: [1805.11525](https://arxiv.org/abs/1805.11525) [[astro-ph.CO](#)].
- [206] Ue-Li Pen et al. “Cosmic Tides”. In: (Feb. 2012). arXiv: [1202.5804](https://arxiv.org/abs/1202.5804) [[astro-ph.CO](#)].
- [207] Ashley Perko et al. “Biased Tracers in Redshift Space in the EFT of Large-Scale Structure”. In: (2016). arXiv: [1610.09321](https://arxiv.org/abs/1610.09321) [[astro-ph.CO](#)].
- [208] Planck Collaboration. “Planck 2013 results. XVIII. The gravitational lensing-infrared background correlation”. In: 571, A18 (2014), A18. DOI: [10.1051/0004-6361/201321540](https://doi.org/10.1051/0004-6361/201321540). arXiv: [1303.5078](https://arxiv.org/abs/1303.5078) [[astro-ph.CO](#)].
- [209] Planck Collaboration. “Planck 2013 results. XVIII. The gravitational lensing-infrared background correlation”. In: *A&A* 571 (2014), A18. DOI: [10.1051/0004-6361/201321540](https://doi.org/10.1051/0004-6361/201321540). URL: <https://doi.org/10.1051/0004-6361/201321540>.
- [210] Planck Collaboration. “Planck 2015 results. X. Diffuse component separation: Foreground maps”. In: 594, A10 (Sept. 2016), A10. DOI: [10.1051/0004-6361/201525967](https://doi.org/10.1051/0004-6361/201525967). arXiv: [1502.01588](https://arxiv.org/abs/1502.01588) [[astro-ph.CO](#)].
- [211] Planck Collaboration. “Planck 2015 results. XXIV. Cosmology from Sunyaev-Zeldovich cluster counts”. In: 594, A24 (2016), A24. DOI: [10.1051/0004-6361/201525833](https://doi.org/10.1051/0004-6361/201525833). arXiv: [1502.01597](https://arxiv.org/abs/1502.01597) [[astro-ph.CO](#)].
- [212] Planck Collaboration. “Planck 2018 results. VIII. Gravitational lensing”. In: *arXiv e-prints*, arXiv:1807.06210 (2018), arXiv:1807.06210. arXiv: [1807.06210](https://arxiv.org/abs/1807.06210) [[astro-ph.CO](#)].
- [213] Planck Collaboration. “Planck intermediate results. XLVIII. Disentangling Galactic dust emission and cosmic infrared background anisotropies”. In: 596, A109 (2016), A109. DOI: [10.1051/0004-6361/201629022](https://doi.org/10.1051/0004-6361/201629022). arXiv: [1605.09387](https://arxiv.org/abs/1605.09387) [[astro-ph.CO](#)].
- [214] J. Prat et al. “Cosmological lensing ratios with DES Y1, SPT, and Planck”. In: 487.1 (2019), pp. 1363–1379. DOI: [10.1093/mnras/stz1309](https://doi.org/10.1093/mnras/stz1309). arXiv: [1810.02212](https://arxiv.org/abs/1810.02212) [[astro-ph.CO](#)].
- [215] William H. Press and Paul Schechter. “Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation”. In: 187 (Feb. 1974), pp. 425–438. DOI: [10.1086/152650](https://doi.org/10.1086/152650).

- [216] A. M. Price-Whelan et al. “The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package”. In: 156, 123 (Sept. 2018), p. 123. DOI: [10.3847/1538-3881/aabc4f](https://doi.org/10.3847/1538-3881/aabc4f).
- [217] Anthony R. Pullen et al. “Constraining gravity at the largest scales through CMB lensing and galaxy velocities”. In: 460.4 (2016), pp. 4098–4108. DOI: [10.1093/mnras/stw1249](https://doi.org/10.1093/mnras/stw1249). arXiv: [1511.04457](https://arxiv.org/abs/1511.04457) [[astro-ph.CO](#)].
- [218] Roland de Putter. “Primordial physics from large-scale structure beyond the power spectrum”. In: (2018). arXiv: [1802.06762](https://arxiv.org/abs/1802.06762) [[astro-ph.CO](#)].
- [219] Roland de Putter, Oliver Zahn, and Eric V. Linder. “CMB lensing constraints on neutrinos and dark energy”. In: *Physical Review D* 79.6 (2009). ISSN: 1550-2368. DOI: [10.1103/PhysRevD.79.065033](https://doi.org/10.1103/PhysRevD.79.065033). URL: <http://dx.doi.org/10.1103/PhysRevD.79.065033>.
- [220] S. Raghunathan et al. “Detection of CMB-Cluster Lensing using Polarization Data from SPTpol”. In: 123.18, 181301 (2019), p. 181301. DOI: [10.1103/PhysRevLett.123.181301](https://doi.org/10.1103/PhysRevLett.123.181301). arXiv: [1907.08605](https://arxiv.org/abs/1907.08605) [[astro-ph.CO](#)].
- [221] S. Raghunathan et al. “Mass Calibration of Optically Selected DES Clusters Using a Measurement of CMB-cluster Lensing with SPTpol Data”. In: 872.2, 170 (Feb. 2019), p. 170. DOI: [10.3847/1538-4357/ab01ca](https://doi.org/10.3847/1538-4357/ab01ca). arXiv: [1810.10998](https://arxiv.org/abs/1810.10998) [[astro-ph.CO](#)].
- [222] Srinivasan Raghunathan, Federico Bianchini, and Christian L. Reichardt. “Imprints of gravitational lensing in the Planck cosmic microwave background data at the location of WISE  $\times$  SCOS galaxies”. In: 98.4, 043506 (2018), p. 043506. DOI: [10.1103/PhysRevD.98.043506](https://doi.org/10.1103/PhysRevD.98.043506). arXiv: [1710.09770](https://arxiv.org/abs/1710.09770) [[astro-ph.CO](#)].
- [223] Beth Reid et al. “SDSS-III Baryon Oscillation Spectroscopic Survey Data Release 12: galaxy target selection and large-scale structure catalogues”. In: 455.2 (Jan. 2016), pp. 1553–1573. DOI: [10.1093/mnras/stv2382](https://doi.org/10.1093/mnras/stv2382). arXiv: [1509.06529](https://arxiv.org/abs/1509.06529) [[astro-ph.CO](#)].
- [224] Mathieu Remazeilles, Jacques Delabrouille, and Jean-François Cardoso. “CMB and SZ effect separation with constrained Internal Linear Combinations”. In: *Monthly Notices of the Royal Astronomical Society* 410.4 (2010), 2481–2487. ISSN: 0035-8711. DOI: [10.1111/j.1365-2966.2010.17624.x](https://doi.org/10.1111/j.1365-2966.2010.17624.x). URL: <http://dx.doi.org/10.1111/j.1365-2966.2010.17624.x>.
- [225] Ashley J. Ross et al. “The Clustering of Galaxies in SDSS-III DR9 Baryon Oscillation Spectroscopic Survey: Constraints on Primordial Non-Gaussianity”. In: *Mon. Not. Roy. Astron. Soc.* 428 (2013), pp. 1116–1127. DOI: [10.1093/mnras/sts094](https://doi.org/10.1093/mnras/sts094). arXiv: [1208.1491](https://arxiv.org/abs/1208.1491) [[astro-ph.CO](#)].
- [226] Noah Sailer, Emmanuel Schaan, and Simone Ferraro. “Lower bias, lower noise CMB lensing with foreground-hardened estimators”. In: 102.6, 063517 (Sept. 2020), p. 063517. DOI: [10.1103/PhysRevD.102.063517](https://doi.org/10.1103/PhysRevD.102.063517). arXiv: [2007.04325](https://arxiv.org/abs/2007.04325) [[astro-ph.CO](#)].
- [227] Noah Sailer et al. *Optimal multi-frequency weighting for CMB lensing*. 2021. arXiv: [2108.01663](https://arxiv.org/abs/2108.01663) [[astro-ph.CO](#)].
- [228] Emmanuel Schaan and Simone Ferraro. “Foreground-Immune Cosmic Microwave Background Lensing with Shear-Only Reconstruction”. In: 122.18, 181301 (2019), p. 181301. DOI: [10.1103/PhysRevLett.122.181301](https://doi.org/10.1103/PhysRevLett.122.181301). arXiv: [1804.06403](https://arxiv.org/abs/1804.06403) [[astro-ph.CO](#)].

- [229] Emmanuel Schaan and Simone Ferraro. “Foreground-Immune Cosmic Microwave Background Lensing with Shear-Only Reconstruction”. In: 122.18, 181301 (May 2019), p. 181301. DOI: [10.1103/PhysRevLett.122.181301](https://doi.org/10.1103/PhysRevLett.122.181301). arXiv: [1804.06403](https://arxiv.org/abs/1804.06403) [[astro-ph.CO](#)].
- [230] Emmanuel Schaan et al. “Atacama Cosmology Telescope: Combined kinematic and thermal Sunyaev-Zel’dovich measurements from BOSS CMASS and LOWZ halos”. In: *Phys. Rev. D* 103 (6 2021), p. 063513. DOI: [10.1103/PhysRevD.103.063513](https://doi.org/10.1103/PhysRevD.103.063513). URL: <https://link.aps.org/doi/10.1103/PhysRevD.103.063513>.
- [231] Emmanuel Schaan et al. “Looking through the same lens: Shear calibration for LSST, Euclid, and WFIRST with stage 4 CMB lensing”. In: 95.12, 123512 (2017), p. 123512. DOI: [10.1103/PhysRevD.95.123512](https://doi.org/10.1103/PhysRevD.95.123512). arXiv: [1607.01761](https://arxiv.org/abs/1607.01761) [[astro-ph.CO](#)].
- [232] David J. Schlegel et al. “Astro2020 APC White Paper: The MegaMapper: a  $z > 2$  Spectroscopic Instrument for the Study of Inflation and Dark Energy”. In: (2019). arXiv: [1907.11171](https://arxiv.org/abs/1907.11171) [[astro-ph.IM](#)].
- [233] Fabian Schmidt. “Towards a self-consistent halo model for the nonlinear large-scale structure”. In: *Phys. Rev. D* 93.6 (2016), p. 063512. DOI: [10.1103/PhysRevD.93.063512](https://doi.org/10.1103/PhysRevD.93.063512). arXiv: [1511.02231](https://arxiv.org/abs/1511.02231) [[astro-ph.CO](#)].
- [234] Fabian Schmidt and Marc Kamionkowski. “Halo Clustering with Non-Local Non-Gaussianity”. In: *Phys. Rev. D* 82 (2010), p. 103002. DOI: [10.1103/PhysRevD.82.103002](https://doi.org/10.1103/PhysRevD.82.103002). arXiv: [1008.0638](https://arxiv.org/abs/1008.0638) [[astro-ph.CO](#)].
- [235] Fabian Schmidt, Enrico Pajer, and Matias Zaldarriaga. “Large-scale structure and gravitational waves. III. Tidal effects”. In: *Physical Review D* 89.8 (2014). ISSN: 1550-2368. DOI: [10.1103/physrevd.89.083507](https://doi.org/10.1103/physrevd.89.083507). URL: <http://dx.doi.org/10.1103/PhysRevD.89.083507>.
- [236] Marcel Schmittfull, Tobias Baldauf, and Uroš Seljak. “Near optimal bispectrum estimators for large-scale structure”. In: *Phys. Rev. D* 91.4 (2015), p. 043530. DOI: [10.1103/PhysRevD.91.043530](https://doi.org/10.1103/PhysRevD.91.043530). arXiv: [1411.6595](https://arxiv.org/abs/1411.6595) [[astro-ph.CO](#)].
- [237] Marcel Schmittfull and Uros Seljak. “Parameter constraints from cross-correlation of CMB lensing with galaxy clustering”. In: *Phys. Rev. D* 97.12 (2018), p. 123540. DOI: [10.1103/PhysRevD.97.123540](https://doi.org/10.1103/PhysRevD.97.123540). arXiv: [1710.09465](https://arxiv.org/abs/1710.09465) [[astro-ph.CO](#)].
- [238] Marcel Schmittfull et al. “Modeling Biased Tracers at the Field Level”. In: *Phys. Rev. D* 100.4 (2019), p. 043514. DOI: [10.1103/PhysRevD.100.043514](https://doi.org/10.1103/PhysRevD.100.043514). arXiv: [1811.10640](https://arxiv.org/abs/1811.10640) [[astro-ph.CO](#)].
- [239] Marcel M. Schmittfull et al. “Joint analysis of CMB temperature and lensing-reconstruction power spectra”. In: 88.6, 063012 (Sept. 2013), p. 063012. DOI: [10.1103/PhysRevD.88.063012](https://doi.org/10.1103/PhysRevD.88.063012). arXiv: [1308.0286](https://arxiv.org/abs/1308.0286) [[astro-ph.CO](#)].
- [240] Roman Scoccimarro, Emiliano Sefusatti, and Matias Zaldarriaga. “Probing primordial non-Gaussianity with large - scale structure”. In: *Phys. Rev. D* 69 (2004), p. 103513. DOI: [10.1103/PhysRevD.69.103513](https://doi.org/10.1103/PhysRevD.69.103513). arXiv: [astro-ph/0312286](https://arxiv.org/abs/astro-ph/0312286).
- [241] Roman Scoccimarro et al. “Large-scale Bias and Efficient Generation of Initial Conditions for Non-Local Primordial Non-Gaussianity”. In: *Phys. Rev. D* 85 (2012), p. 083002. DOI: [10.1103/PhysRevD.85.083002](https://doi.org/10.1103/PhysRevD.85.083002). arXiv: [1108.5512](https://arxiv.org/abs/1108.5512) [[astro-ph.CO](#)].

- [242] Roman Scoccimarro et al. “Nonlinear Evolution of the Bispectrum of Cosmological Perturbations”. In: *The Astrophysical Journal* 496.2 (1998), 586–604. ISSN: 1538-4357. DOI: [10.1086/305399](https://doi.org/10.1086/305399). URL: <http://dx.doi.org/10.1086/305399>.
- [243] Emiliano Sefusatti et al. “Cosmology and the Bispectrum”. In: *Phys. Rev. D* 74 (2006), p. 023522. DOI: [10.1103/PhysRevD.74.023522](https://doi.org/10.1103/PhysRevD.74.023522). arXiv: [astro-ph/0604505](https://arxiv.org/abs/astro-ph/0604505).
- [244] Neelima Sehgal et al. “Simulations of the Microwave Sky”. In: 709.2 (Feb. 2010), pp. 920–936. DOI: [10.1088/0004-637X/709/2/920](https://doi.org/10.1088/0004-637X/709/2/920). arXiv: [0908.0540](https://arxiv.org/abs/0908.0540) [[astro-ph.CO](https://arxiv.org/abs/0908.0540)].
- [245] Neelima Sehgal et al. “Simulations of the Microwave Sky”. In: 709.2 (2010), pp. 920–936. DOI: [10.1088/0004-637X/709/2/920](https://doi.org/10.1088/0004-637X/709/2/920). arXiv: [0908.0540](https://arxiv.org/abs/0908.0540) [[astro-ph.CO](https://arxiv.org/abs/0908.0540)].
- [246] Uros Seljak. “Extracting primordial non-gaussianity without cosmic variance”. In: *Phys. Rev. Lett.* 102 (2009), p. 021302. DOI: [10.1103/PhysRevLett.102.021302](https://doi.org/10.1103/PhysRevLett.102.021302). arXiv: [0807.1770](https://arxiv.org/abs/0807.1770) [[astro-ph](https://arxiv.org/abs/0807.1770)].
- [247] Uroš Seljak, Nico Hamaus, and Vincent Desjacques. “How to Suppress the Shot Noise in Galaxy Surveys”. In: 103.9, 091303 (Aug. 2009), p. 091303. DOI: [10.1103/PhysRevLett.103.091303](https://doi.org/10.1103/PhysRevLett.103.091303). arXiv: [0904.2963](https://arxiv.org/abs/0904.2963) [[astro-ph.CO](https://arxiv.org/abs/0904.2963)].
- [248] Uros Seljak and Matias Zaldarriaga. “A Line of Sight Approach to Cosmic Microwave Background Anisotropies”. In: *The Astrophysical Journal* 469 (1996), p. 437. DOI: [10.1086/177793](https://doi.org/10.1086/177793). arXiv: [9603033](https://arxiv.org/abs/9603033) [[astro-ph](https://arxiv.org/abs/9603033)]. URL: <http://arxiv.org/abs/astro-ph/9603033><http://dx.doi.org/10.1086/177793>.
- [249] Leonardo Senatore. *Lectures on inflation*.
- [250] Cien Shang et al. “Improved models for cosmic infrared background anisotropies: new constraints on the infrared galaxy population”. In: *Monthly Notices of the Royal Astronomical Society* 421.4 (2012), 2832–2845. ISSN: 0035-8711. DOI: [10.1111/j.1365-2966.2012.20510.x](https://doi.org/10.1111/j.1365-2966.2012.20510.x). URL: <http://dx.doi.org/10.1111/j.1365-2966.2012.20510.x>.
- [251] J. Richard Shaw et al. “Coaxing cosmic 21 cm fluctuations from the polarized sky using m-mode analysis”. In: *Phys. Rev. D* 91.8 (2015), p. 083514. DOI: [10.1103/PhysRevD.91.083514](https://doi.org/10.1103/PhysRevD.91.083514). arXiv: [1401.2095](https://arxiv.org/abs/1401.2095) [[astro-ph.CO](https://arxiv.org/abs/1401.2095)].
- [252] Blake D. Sherwin and Matias Zaldarriaga. “The Shift of the Baryon Acoustic Oscillation Scale: A Simple Physical Picture”. In: *Phys. Rev. D* 85 (2012), p. 103523. DOI: [10.1103/PhysRevD.85.103523](https://doi.org/10.1103/PhysRevD.85.103523). arXiv: [1202.3998](https://arxiv.org/abs/1202.3998) [[astro-ph.CO](https://arxiv.org/abs/1202.3998)].
- [253] Blake D. Sherwin et al. “Evidence for Dark Energy from the Cosmic Microwave Background Alone Using the Atacama Cosmology Telescope Lensing Measurements”. In: *Physical Review Letters* 107.2 (2011). ISSN: 1079-7114. DOI: [10.1103/physrevlett.107.021302](https://doi.org/10.1103/physrevlett.107.021302). URL: <http://dx.doi.org/10.1103/PhysRevLett.107.021302>.
- [254] Blake D. Sherwin et al. “Evidence for Dark Energy from the Cosmic Microwave Background Alone Using the Atacama Cosmology Telescope Lensing Measurements”. In: 107.2, 021302 (2011), p. 021302. DOI: [10.1103/PhysRevLett.107.021302](https://doi.org/10.1103/PhysRevLett.107.021302). arXiv: [1105.0419](https://arxiv.org/abs/1105.0419) [[astro-ph.CO](https://arxiv.org/abs/1105.0419)].
- [255] Blake D. Sherwin et al. “The Atacama Cosmology Telescope: Cross-correlation of cosmic microwave background lensing and quasars”. In: 86.8, 083006 (2012), p. 083006. DOI: [10.1103/PhysRevD.86.083006](https://doi.org/10.1103/PhysRevD.86.083006). arXiv: [1207.4543](https://arxiv.org/abs/1207.4543) [[astro-ph.CO](https://arxiv.org/abs/1207.4543)].

- [256] Blake D. Sherwin et al. “Two-season Atacama Cosmology Telescope polarimeter lensing power spectrum”. In: 95.12, 123529 (2017), p. 123529. DOI: [10.1103/PhysRevD.95.123529](https://doi.org/10.1103/PhysRevD.95.123529). arXiv: [1611.09753](https://arxiv.org/abs/1611.09753) [[astro-ph.CO](#)].
- [257] Ravi K. Sheth and Giuseppe Tormen. “Large-scale bias and the peak background split”. In: *Monthly Notices of the Royal Astronomical Society* 308.1 (Sept. 1999), pp. 119–126. ISSN: 0035-8711. DOI: [10.1046/j.1365-8711.1999.02692.x](https://doi.org/10.1046/j.1365-8711.1999.02692.x). eprint: <https://academic.oup.com/mnras/article-pdf/308/1/119/18409158/308-1-119.pdf>. URL: <https://doi.org/10.1046/j.1365-8711.1999.02692.x>.
- [258] Joseph Silk. “Cosmic Black-Body Radiation and Galaxy Formation”. In: 151 (Feb. 1968), p. 459. DOI: [10.1086/149449](https://doi.org/10.1086/149449).
- [259] Anže Slosar et al. “Scratches from the Past: Inflationary Archaeology through Features in the Power Spectrum of Primordial Fluctuations”. In: (Mar. 2019). arXiv: [1903.09883](https://arxiv.org/abs/1903.09883) [[astro-ph.CO](#)].
- [260] Anže Slosar et al. “Constraints on local primordial non-Gaussianity from large scale structure”. In: *Journal of Cosmology and Astroparticle Physics* 2008.08 (2008), p. 031. ISSN: 1475-7516. DOI: [10.1088/1475-7516/2008/08/031](https://doi.org/10.1088/1475-7516/2008/08/031). URL: <http://dx.doi.org/10.1088/1475-7516/2008/08/031>.
- [261] Anže Slosar et al. “Packed Ultra-wideband Mapping Array (PUMA): A Radio Telescope for Cosmology and Transients”. In: (2019). arXiv: [1907.12559](https://arxiv.org/abs/1907.12559) [[astro-ph.IM](#)].
- [262] Kendrick M. Smith, Oliver Zahn, and Olivier Doré. “Detection of gravitational lensing in the cosmic microwave background”. In: *Physical Review D - Particles, Fields, Gravitation and Cosmology* 76.4 (2007), p. 043510. ISSN: 15507998. DOI: [10.1103/PhysRevD.76.043510](https://doi.org/10.1103/PhysRevD.76.043510). URL: <https://journals.aps.org/prd/abstract/10.1103/PhysRevD.76.043510>.
- [263] Kendrick M. Smith, Oliver Zahn, and Olivier Doré. “Detection of gravitational lensing in the cosmic microwave background”. In: 76.4, 043510 (2007), p. 043510. DOI: [10.1103/PhysRevD.76.043510](https://doi.org/10.1103/PhysRevD.76.043510). arXiv: [0705.3980](https://arxiv.org/abs/0705.3980) [[astro-ph](#)].
- [264] Kendrick M. Smith et al. “Gravitational Lensing”. In: *AIP Conference Proceedings* 1141.1 (2009), pp. 121–178. DOI: [10.1063/1.3160886](https://doi.org/10.1063/1.3160886). eprint: <https://aip.scitation.org/doi/pdf/10.1063/1.3160886>. URL: <https://aip.scitation.org/doi/abs/10.1063/1.3160886>.
- [265] Yong-Seon Song et al. “The Far-Infrared Background Correlation with Cosmic Microwave Background Lensing”. In: *The Astrophysical Journal* 590.2 (2003), pp. 664–672. DOI: [10.1086/375188](https://doi.org/10.1086/375188). URL: <https://doi.org/10.1086%2F375188>.
- [266] Volker Springel. “The Cosmological simulation code GADGET-2”. In: *Mon. Not. Roy. Astron. Soc.* 364 (2005), pp. 1105–1134. DOI: [10.1111/j.1365-2966.2005.09655.x](https://doi.org/10.1111/j.1365-2966.2005.09655.x). arXiv: [astro-ph/0505010](https://arxiv.org/abs/astro-ph/0505010).
- [267] George Stein et al. “The Websky Extragalactic CMB Simulations”. In: *arXiv e-prints*, arXiv:2001.08787 (2020), arXiv:2001.08787. arXiv: [2001.08787](https://arxiv.org/abs/2001.08787) [[astro-ph.CO](#)].
- [268] George Stein et al. “The Websky extragalactic CMB simulations”. In: *Journal of Cosmology and Astroparticle Physics* 2020.10 (2020), 012â012. ISSN: 1475-7516. DOI: [10.1088/1475-7516/2020/10/012](https://doi.org/10.1088/1475-7516/2020/10/012). URL: <http://dx.doi.org/10.1088/1475-7516/2020/10/012>.

- [269] K. T. Story et al. “A MEASUREMENT OF THE COSMIC MICROWAVE BACKGROUND DAMPING TAIL FROM THE 2500-SQUARE-DEGREE SPT-SZ SURVEY”. In: *The Astrophysical Journal* 779.1 (2013), p. 86. ISSN: 0004-637X. DOI: [10.1088/0004-637X/779/1/86](https://doi.org/10.1088/0004-637X/779/1/86).
- [270] K. T. Story et al. “A Measurement of the Cosmic Microwave Background Gravitational Lensing Potential from 100 Square Degrees of SPTpol Data”. In: 810.1, 50 (2015), p. 50. DOI: [10.1088/0004-637X/810/1/50](https://doi.org/10.1088/0004-637X/810/1/50). arXiv: [1412.4760 \[astro-ph.CO\]](https://arxiv.org/abs/1412.4760).
- [271] Naonori S. Sugiyama et al. “Perturbation theory approach to predict the covariance matrices of the galaxy power spectrum and bispectrum in redshift space”. In: (Aug. 2019). arXiv: [1908.06234 \[astro-ph.CO\]](https://arxiv.org/abs/1908.06234).
- [272] R. A. Sunyaev and Ya. B. Zeldovich. “Small-Scale Fluctuations of Relic Radiation”. In: 7.1 (Apr. 1970), pp. 3–19. DOI: [10.1007/BF00653471](https://doi.org/10.1007/BF00653471).
- [273] Max Tegmark, Andy Taylor, and Alan Heavens. “Karhunen-Loeve eigenvalue problems in cosmology: How should we tackle large data sets?” In: *Astrophys. J.* 480 (1997), p. 22. DOI: [10.1086/303939](https://doi.org/10.1086/303939). arXiv: [astro-ph/9603021](https://arxiv.org/abs/astro-ph/9603021).
- [274] Matteo Tellarini et al. “Non-local bias in the halo bispectrum with primordial non-Gaussianity”. In: *JCAP* 1507.07 (2015), p. 004. DOI: [10.1088/1475-7516/2015/07/004](https://doi.org/10.1088/1475-7516/2015/07/004). arXiv: [1504.00324 \[astro-ph.CO\]](https://arxiv.org/abs/1504.00324).
- [275] R. J. Thornton et al. “The Atacama Cosmology Telescope: The Polarization-sensitive ACTPol Instrument”. In: 227.2, 21 (Dec. 2016), p. 21. DOI: [10.3847/1538-4365/227/2/21](https://doi.org/10.3847/1538-4365/227/2/21). arXiv: [1605.06569 \[astro-ph.IM\]](https://arxiv.org/abs/1605.06569).
- [276] Alberto Vallinotto. “Using Cosmic Microwave Background Lensing to Constrain the Multiplicative Bias of Cosmic Shear”. In: 759.1, 32 (2012), p. 32. DOI: [10.1088/0004-637X/759/1/32](https://doi.org/10.1088/0004-637X/759/1/32). arXiv: [1110.5339 \[astro-ph.CO\]](https://arxiv.org/abs/1110.5339).
- [277] A. van Engelen et al. “CMB Lensing Power Spectrum Biases from Galaxies and Clusters Using High-angular Resolution Temperature Maps”. In: 786.1, 13 (2014), p. 13. DOI: [10.1088/0004-637X/786/1/13](https://doi.org/10.1088/0004-637X/786/1/13). arXiv: [1310.7023 \[astro-ph.CO\]](https://arxiv.org/abs/1310.7023).
- [278] A. van Engelen et al. “CMB Lensing Power Spectrum Biases from Galaxies and Clusters Using High-angular Resolution Temperature Maps”. In: 786.1, 13 (May 2014), p. 13. DOI: [10.1088/0004-637X/786/1/13](https://doi.org/10.1088/0004-637X/786/1/13). arXiv: [1310.7023 \[astro-ph.CO\]](https://arxiv.org/abs/1310.7023).
- [279] Alexander van Engelen et al. “The Atacama Cosmology Telescope: Lensing of CMB Temperature and Polarization Derived from Cosmic Infrared Background Cross-correlation”. In: 808.1, 7 (2015), p. 7. DOI: [10.1088/0004-637X/808/1/7](https://doi.org/10.1088/0004-637X/808/1/7). arXiv: [1412.0626 \[astro-ph.CO\]](https://arxiv.org/abs/1412.0626).
- [280] Flavien Vansyngel et al. “Semi-blind Bayesian inference of CMB map and power spectrum”. In: *Astronomy Astrophysics* 588 (2016), A113. ISSN: 1432-0746. DOI: [10.1051/0004-6361/201424890](https://doi.org/10.1051/0004-6361/201424890). URL: <http://dx.doi.org/10.1051/0004-6361/201424890>.
- [281] Zvonimir Vlah, Emanuele Castorina, and Martin White. “The Gaussian streaming model and convolution Lagrangian effective field theory”. In: *Journal of Cosmology and Astroparticle Physics* 2016.12 (2016), pp. 007–007. DOI: [10.1088/1475-7516/2016/12/007](https://doi.org/10.1088/1475-7516/2016/12/007). URL: <https://doi.org/10.1088/1475-7516/2016/12/007>.

- [282] Woldemar Voigt. *Lehrbuch der kristallphysik (mit ausschluss der kristalloptik)*. Leipzig; Berlin: B.G. Teubner, 1910.
- [283] Steven Weinberg. “Adiabatic modes in cosmology”. In: *Physical Review D* 67.12 (2003). ISSN: 1089-4918. DOI: [10.1103/physrevd.67.123504](https://doi.org/10.1103/physrevd.67.123504). URL: <http://dx.doi.org/10.1103/PhysRevD.67.123504>.
- [284] Steven Weinberg. *Cosmology*. Oxford University, 2008.
- [285] Martin White, Jeremy L. Tinker, and Cameron K. McBride. “Mock galaxy catalogues using the quick particle mesh method”. In: 437.3 (Jan. 2014), pp. 2594–2606. DOI: [10.1093/mnras/stt2071](https://doi.org/10.1093/mnras/stt2071). arXiv: [1309.5532](https://arxiv.org/abs/1309.5532) [[astro-ph.CO](https://arxiv.org/abs/1309.5532)].
- [286] Martin White et al. “Interferometric Observation of Cosmic Microwave Background Anisotropies”. In: *The Astrophysical Journal* 514.1 (1999), 12–24. ISSN: 1538-4357. DOI: [10.1086/306911](https://doi.org/10.1086/306911). URL: <http://dx.doi.org/10.1086/306911>.
- [287] S. D. M. White and M. J. Rees. “Core condensation in heavy halos: a two-stage theory for galaxy formation and clustering.” In: 183 (May 1978), pp. 341–358. DOI: [10.1093/mnras/183.3.341](https://doi.org/10.1093/mnras/183.3.341).
- [288] Michael J. Wilson and Martin White. “Cosmology with dropout selection: straw-man surveys & CMB lensing”. In: *JCAP* 1910.10 (2019), p. 015. DOI: [10.1088/1475-7516/2019/10/015](https://doi.org/10.1088/1475-7516/2019/10/015). arXiv: [1904.13378](https://arxiv.org/abs/1904.13378) [[astro-ph.CO](https://arxiv.org/abs/1904.13378)].
- [289] W. L. K. Wu et al. “A Measurement of the Cosmic Microwave Background Lensing Potential and Power Spectrum from 500 deg<sup>2</sup> of SPTpol Temperature and Polarization Data”. In: *The Astrophysical Journal* 884.1 (2019), p. 70. DOI: [10.3847/1538-4357/ab4186](https://doi.org/10.3847/1538-4357/ab4186). URL: <https://doi.org/10.3847/1538-4357/ab4186>.
- [290] Kohji Yoshikawa et al. “Nonlinear Stochastic Biasing of Galaxies and Dark Halos in Cosmological Hydrodynamic Simulations”. In: *The Astrophysical Journal* 558.2 (2001), pp. 520–534. DOI: [10.1086/322445](https://doi.org/10.1086/322445). URL: <https://doi.org/10.1086/322445>.
- [291] Matias Zaldarriaga and Uroš Seljak. “Reconstructing projected matter density power spectrum from cosmic microwave background”. In: 59.12, 123507 (1999), p. 123507. DOI: [10.1103/PhysRevD.59.123507](https://doi.org/10.1103/PhysRevD.59.123507). arXiv: [astro-ph/9810257](https://arxiv.org/abs/astro-ph/9810257) [[astro-ph](https://arxiv.org/abs/astro-ph/9810257)].
- [292] Hong-Ming Zhu et al. “Cosmic tidal reconstruction”. In: *Phys. Rev. D* 93.10 (2016), p. 103504. DOI: [10.1103/PhysRevD.93.103504](https://doi.org/10.1103/PhysRevD.93.103504). arXiv: [1511.04680](https://arxiv.org/abs/1511.04680) [[astro-ph.CO](https://arxiv.org/abs/1511.04680)].
- [293] Hong-Ming Zhu et al. “Recovering lost 21 cm radial modes via cosmic tidal reconstruction”. In: *Phys. Rev. D* 98.4 (2018), p. 043511. DOI: [10.1103/PhysRevD.98.043511](https://doi.org/10.1103/PhysRevD.98.043511). arXiv: [1610.07062](https://arxiv.org/abs/1610.07062) [[astro-ph.CO](https://arxiv.org/abs/1610.07062)].
- [294] Andrea Zonca et al. “healpy: equal area pixelization and spherical harmonics transforms for data on the sphere in Python”. In: *Journal of Open Source Software* 4.35 (Mar. 2019), p. 1298. DOI: [10.21105/joss.01298](https://doi.org/10.21105/joss.01298). URL: <https://doi.org/10.21105/joss.01298>.
- [295] Íñigo Zubeldia and Anthony Challinor. “Cosmological constraints from Planck galaxy clusters with CMB lensing mass bias calibration”. In: 489.1 (2019), pp. 401–419. DOI: [10.1093/mnras/stz2153](https://doi.org/10.1093/mnras/stz2153). arXiv: [1904.07887](https://arxiv.org/abs/1904.07887) [[astro-ph.CO](https://arxiv.org/abs/1904.07887)].