THE UNIVERSITY OF CHICAGO

FIELD-LEVEL COSMOLOGICAL INFERENCE

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF PHYSICS

BY
JUNZHE ZHOU

CHICAGO, ILLINOIS
MARCH 2026

Copyright © 2026 by JUNZHE ZHOU

All Rights Reserved

To my family, and especially my mother, for twenty-five years of kindness and love.

TABLE OF CONTENTS

LIST OF FIGURES viii

LIST OF TABLES xxiii

ACKNOWLEDGMENTS xxv

1 INTRODUCTION 1

2 FIELD-LEVEL MULTIPROBE ANALYSIS OF THE CMB, INTEGRATED SACHS-WOLFE EFFECT, AND THE GALAXY DENSITY MAPS . . . 9
- 2.1 Abstract 9
- 2.2 Introduction 9
- 2.3 Theoretical framework 15
  - 2.3.1 Field-level multiprobe analysis 15
  - 2.3.2 Pixels 18
  - 2.3.3 Methodology 19
- 2.4 Reconstruction of the primordial CMB map 23
  - 2.4.1 Biases of the optimal estimator 24
  - 2.4.2 Sampling the CMB field 28
- 2.5 Joint reconstruction of the primordial CMB and the ISW effect . . . 34
  - 2.5.1 The ISW effect 34
  - 2.5.2 Separating the primordial and ISW signals 36
- 2.6 Joint reconstruction of the CMB, ISW, and the galaxy density maps 45
  - 2.6.1 Theoretical covariance 46
  - 2.6.2 Cosmological parameters 48
  - 2.6.3 Noise model and simulation 50
  - 2.6.4 Posterior modeling and sampling 52
  - 2.6.5 Results 53
- 2.7 Conclusions 58
  - 2.7.1 General cosmological parameters 60
  - 2.7.2 Masking, anisotropic noise, and other systematic effects 61
  - 2.7.3 Smaller scales 62
- 2.8 Appendix 63
  - 2.8.1 Distributions of individual modes of the sampled power spectra 63
  - 2.8.2 MAP and sampler estimators 63
  - 2.8.3 MAP estimator 68
  - 2.8.4 Chain convergence 70

3 ACCURATE FIELD-LEVEL WEAK LENSING INFERENCE FOR PRECISION COSMOLOGY . . . 72
3.1 Abstract . . . 72
3.2 Introduction . . . 72
3.3 Weak lensing basics . . . 75
3.3.1 Weak lensing maps . . . 75
3.3.2 Weak lensing statistics . . . 76
3.3.3 Weak lensing observables . . . 79
3.4 Simulations and Datasets . . . 80
3.5 Inference Framework . . . 82
3.6 Forward modeling and analysis systematics . . . 86
3.6.1 Map making . . . 86
3.6.2 Smoothing and aliasing . . . 87
3.6.3 Boundary conditions . . . 93
3.6.4 Fourier mode coupling . . . 94
3.6.5 Number density-induced shape noise . . . 96
3.7 Model misspecification . . . 98
3.7.1 Impact on cosmological parameters . . . 100
3.7.2 Map reconstructions . . . 111
3.8 Summary . . . 115
3.9 Appendix . . . 118
3.10 Caliberation of the Takahashi simulations' power spectra . . . 118

4 A HAMILTONIAN, POST-BORN, THREE-DIMENSIONAL, ON-THE-FLY RAY TRACING ALGORITHM FOR GRAVITATIONAL LENSING . . . 119
4.1 Abstract . . . 119
4.2 Introduction . . . 120
4.3 Hamiltonian dynamics of light rays . . . 124
4.4 Symplectic ray tracing . . . 127
4.4.1 Equations of motion in discrete time . . . 127
4.4.2 Time resolution correction . . . 129
4.5 Weak lensing observables . . . 129
4.6 Implementation . . . 133
4.6.1 Efficient dark matter and photons co-evolution . . . 133
4.6.2 The kick operator . . . 135
4.6.3 Adaptive ray mesh . . . 136
4.7 Validations . . . 137
4.7.1 Lensing by a point mass . . . 137

4.7.2 Post-Born weak lensing in a cosmological volume 139
4.8 Conclusions 146
4.9 Appendix 148
4.9.1 Hamiltonian dynamics of photons lensed by weak gravitational field 148
4.9.2 Symplectic integrator 153
4.9.3 Adaptive ray mesh 154
4.9.4 Other considerations and convergence tests for point mass lensing 156
4.9.5 Convergence tests for weak lensing power spectrum 158

5 MAP-LEVEL BARYONIFICATION: UNIFIED TREATMENT OF WEAK LENSING TWO-POINT AND HIGHER-ORDER STATISTICS 160
5.1 Abstract 160
5.2 Introduction 161
5.3 Baryonification of simulations with BaryonForge 165
5.4 Simulations 169
5.4.1 FLAMINGO simulations 169
5.4.2 DES-like weak lensing convergence maps 173
5.5 Statistics 174
5.5.1 Power Spectrum 175
5.5.2 Scattering Transform Coefficients (ST) 177
5.5.3 Wavelet Phase Harmonics Coefficients (WPH) 179
5.5.4 Higher Order Moments 180
5.5.5 The statistics of FLAMINGO and baryonified convergence maps 181
5.6 Method 183
5.6.1 Emulator 185
5.6.2 Sampling 187
5.7 Results 189
5.7.1 The 10-parameter baryon correction model 189
5.7.2 Baryon correction model with restricted degrees of freedoms . 191
5.7.3 Baryon parameters 193
5.8 Conclusion 196
5.9 Appendix 200
5.9.1 Scale cut 200
5.9.2 Statistics as a function of baryon parameters 200
5.9.3 Emulator 202
5.9.4 Best-fit baryon parameters 204

6 THE ISOTROPY OF HUBBLE EXPANSION IN THE EARLY AND LATE UNIVERSE 210
6.1 Abstract 210
6.2 Introduction 210
6.3 Hubble residual maps 212
6.4 Analysis 219
6.5 Conclusion 222

7 FIELD-LEVEL DE-CONTAMINATION OF ATMOSPHERIC FOREGROUNDS FOR GROUND-BASED CMB EXPERIMENTS 225
7.1 Introduction 225
7.2 Forward model 227
7.2.1 Time-ordered data model 227
7.2.2 Atmospheric emission modeling 229
7.3 A two-pixel example 230
7.4 Joint regression for CMB and per-scan frozen atmosphere 232
7.4.1 Global optimal CMB reconstruction 233
7.4.2 Per-scan CMB reconstruction and optimal synthesis 235
7.5 The geometry and properties of the optimal estimators 236
7.5.1 The large detector noise limit 237
7.5.2 The large atmosphere emission amplitude limit 238
7.6 Two dimensional model and multi-scan synthesis 244
7.7 Appendix 247
7.7.1 Numerical implementation 247
7.7.2 Marginalization over atmospheric screens 248
7.7.3 The equivalence between global MAP estimator and the synthesized per-scan solutions 249

REFERENCES 251

LIST OF FIGURES

2.1 The flow chart of a general field-level multiprobe analysis that accumulates samples of both the cosmological parameters and the tracer maps (values of each signal in each pixel) as discussed in §2.3.1 and §2.3.3. We start from the prior distribution of the cosmological and latent map parameters. We then use the realized cosmological parameters to construct the covariance of the tracers, which in turn transforms the latent map parameters into physical tracer maps. The covariance, tracer maps, and observed data are then combined into the likelihood and – after multiplying by the priors – the posterior functions. If the above calculations are all programmatically differentiable, we can calculate the derivatives of the posterior function easily, and use HMC No-U-Turn Sampler (HMC-NUTS) to efficiently sample from the very high dimensional posterior space. In this diagram, diamonds denote sampled parameters, squares denote model-relevant functions, and the hexagon is the (fixed) observed data vector. The pink diamond represent samples that represents the posterior space. 17

2.2 A numerical experiment of CMB reconstruction that demonstrates the difference between the MAP estimators and the sampling-based estimators. We generated the truth map from a fiducial $\mathbb{C}^{\mathcal{P}}$ and added isotropic noise. The truth and noise spectra are shown in black and purple. We test two reconstruction models, the first has a fixed $\mathbb{C}^{\mathcal{P}}$ (Eq. 2.11), and the second has a free amplitude A^2 modulating the fiducial $\mathbb{C}^{\mathcal{P}}$. Their MAP solutions are given by Eqs. 2.13 and 2.15, and are shown in green and red respectively. We confirm the analytical solutions match the results from direct numerical optimization. Both two-point and MAP estimators are biased lower than truth and the free amplitude model has a greater bias. We then sampled $s^{\mathcal{P}}$ directly from the posterior space. The spectra of the mean maps for the fixed and free amplitude cases are biased (both overlap the green curve). However, in both cases, the distributions (the orange shaded region represent 1σ credible interval) of the spectra scatter around the truth and are unbiased. 27

2.3 The scatter plot of the ratio between the real components of the MAP a_{lm} and that of the truth a_{lm} as a function of l. The cases for fixed and varying amplitude parameters are shown in the upper and lower panels respectively. In each case, the mean of the ratios (black curve) matches the Wiener filter expectations (with the MAP amplitude in the case of varying amplitude). The color map represents point cloud density. 29

2.4 The posterior distribution of the amplitude A (blue curve) compared to truth (black dashed line). The shaded orange area represents the Fisher forecast of the 1σ uncertainty centered on truth, and the shaded blue area represents the 1σ uncertainty reported by the sampler. The red line shows the MAP value for A, which is quite far from truth. 33
2.5 The power spectra of the primordial CMB and the ISW effect on large scales (low-ℓ modes). 37
2.6 The recovered primordial ($s^{\mathcal{P}}$) and ISW ($s^{\mathcal{I}}$) spectra using the various models in §2.5.2. The simulation is drawn from the fiducial cosmology and a noise variance of 200 μK^2. The top and bottom panels show the case of fixed and free amplitude respectively, and the left and right panels show the primordial and the ISW results, respectively. For each case, the truth spectrum is shown in black, the MAP result is shown in green, and the 1σ credible interval of the sampled spectra is shown in shaded orange. The effective noise of the primordial maps is $\mathbb{C}^{\mathcal{I}} + \mathbb{C}^{n,\mathcal{T}}$ shown in purple. The effective noise of the ISW map ($\mathbb{C}^{\mathcal{P}} + \mathbb{C}^{n,\mathcal{T}}$) is above the ISW spectra in the right panels on all scales and hence not shown in the plot. Notice that in the case of free amplitude, the MAP solution is different from the $\mathbb{C}_l$ of the sampled maps (red). 39
2.7 The scatter plot of the ratio between the MAP and the truth a_{lm}'s real components. The upper panel shows the case with fixed amplitude and the bottom shows the case with free amplitude. In the case of the fixed amplitude, the mean of the ratios is again the Wiener filter factor (black). In the case of free amplitude, the MAP solution has a null spectrum. Thus the ratios scatter around 0 for all scales. 40
2.8 Comparison between the observed data (row 1), truth maps, MAP maps, and the mean sampled maps. Rows 2-3 show the model with fixed amplitude, while rows 4-5 show the model of free amplitude. The color bar is shared across each row. Note that in the case of fixed amplitude, the mean sampled map is equal to the MAP map (second and third columns in the second and third rows), while when the amplitude is varied, the two diverge (same columns in the fourth and last row, but most obvious in the ISW maps in the last row) since the posterior marginalized over the amplitude is no longer a simple Gaussian in the signals. 42

2.9 The posterior distribution of the primordial and ISW amplitudes with parameter bounds. The data strongly constrain the sum of the two amplitudes (a fact not obvious here because the x-range is so much smaller than the y-range), but the sampling technique manages to correctly infer the values of both. 43
2.10 The galaxy number densities for each redshift bin as a function of redshift z. $N_i(z)$ is qualitatively modeled on the DES Year 3 MagLim sample. . . 47
2.11 The correlation structure between different tracers and the sampler results. The lower-left portion shows the correlation coefficients (blue) between the ISW and the 6 DES Year3 MagLim -like galaxy tracers from different redshift bins. The primordial field is not shown here since it is independent of all late-time tracers. One crucial observation is that (lowest-left panel) the ISW effect correlates most strongly with the low redshift galaxy density maps, since only at late times did dark energy significantly drive the accelerated cosmic expansion. The upper right corner shows the distribution of power spectra sampled from the joint posterior Eq. 2.37 compared to the true power spectra. The input galaxy and ISW signals are shown in black while the distribution of the posterior samples is shown in orange. The diagonal subplots also contain the noise (auto)spectra. Note that the galaxy autospectrum is unitless since the galaxy redshift distribution is normalized. The ISW autospectrum has the unit of μK^2. Because of the different numeric scales, the galaxy-ISW cross-spectra and the ISW autospectrum are broken off into their own column with independent y-axis limits. 49
2.12 The recovered $s^{\mathcal{P}}$ and $s^{\mathcal{I}}$ spectra using the joint reconstruction technique described in §2.6. For each case, the truth spectrum is shown in black, the 1σ credible interval of the sampled spectra is shown in shaded orange, and the spectrum of the mean map is shown in red. The effective noise of the primordial maps is $\mathbb{C}^{\mathcal{I}} + \mathbb{C}^{n,\mathcal{T}}$ shown in purple. The effective noise of the CMB map ($\mathbb{C}^{\mathcal{P}} + \mathbb{C}^{n,\mathcal{T}}$) is above the ISW spectra on all scales and hence not shown in the plot. Notice that both tracers are reconstructed with much higher SNR comparing to the CMB-ISW model. 51
2.13 The posterior distribution of the amplitudes $A^{\mathcal{P}}$ and $A^{\mathcal{I}}$ and the 6 galaxy biases b_i's. The 2-dimensional contours label the 0.68 and 0.95 credible intervals. The KDE-smoothed histograms above the contours are the parameters' marginal distributions, where the shaded intervals represent the 0.16 and 0.84 quantiles. The truth values are indicated by dashed black lines. 55

2.14 Comparison between the observed temperature (row 1), the primordial CMB maps (row 2), and the ISW maps (row 3). For both the primordial CMB and the ISW maps, we compare the truth, the mean sampled, and the difference maps obtained from the CMB-ISW-galaxy model with free cosmological parameters. 57
2.15 Map-level reconstruction of the galaxy fields. Here we only show the first and the last redshift bins, the results for other bins are similar. Unit is in dimensionless overdensity. Note that although the two redshift bins have the same map resolution, the pixel scale of the lower redshift bin corresponds to smaller physical scale. Thus, the lower redshift bin has higher map-level variance, as expected. 58
2.16 The marginal distribution of $\mathbb{C}^{\mathcal{P}}$ of the posterior samples for the CMB model (§2.4). The case with fixed fiducial power spectrum is shown in orange, and the case where one admits a free amplitude is shown in blue. The vertical black lines show the truth $\mathbb{C}^{\mathcal{P}}$, while the colored vertical lines show the MAP values for $\mathbb{C}^{\mathcal{P}}$. 64
2.17 The marginal distribution of $\mathbb{C}^{\mathcal{I}}$ of the posterior samples for the CMB-ISW model (§2.5). The case with fixed fiducial power spectrum is shown in orange, and the case where one admits a free amplitude is shown in blue. The vertical black lines show the truth $\mathbb{C}^{\mathcal{I}}$, while the colored vertical lines show the MAP values for $\mathbb{C}^{\mathcal{I}}$. 65
2.18 The blue lines show the marginal distribution of $\mathbb{C}^{\mathcal{I}}$ of the posterior samples for the CMB-ISW-galaxy model (§2.6). The black and the red dashed lines mark the truth value and the mean of the sample distribution respectively. 66
2.19 The chain convergence of the power spectra amplitudes and galaxy biases for the CMB-ISW-galaxy model. The left panels show the posterior distribution of each parameter after KDE smoothing. The right panels show each parameter's value as a function of sample index after 3000 warmup steps. 68
2.20 The auto-correlation functions of the 8 sampled cosmological parameters shown in Fig. 2.19. 69

3.1 Examples of Gaussian, log-normal, and Takahashi fields. All three fields have the same two-point statistics (modeled after the first redshift bin § 3.4) but a visibly different field-level appearance. Their one-point PDFs are shown in the right plot, where we see that the log-normal field has a stronger cutoff at small κ compared to the others. We limit the maximum color map scale to $\kappa = 0.06$ to emphasize the visual distinction between the fields. 77
3.2 The normalized PDF of the tomographic redshift distributions of the HSC Year 3 galaxy shape catalog [1]. 81
3.3 This flow chart describes `Miko`'s weak lensing field level forward model and inference pipeline (§ 3.5). The pipeline has four steps (from top to bottom): generating latent parameters, forward modeling the noiseless fields, computing the likelihood given the data, and sampling. In the chart, diamonds represent sampled parameters, squares the model-relevant functions, and the hexagon the observed data vector. 84
3.4 The impact of smoothing and aliasing on C is presented in terms of the fractional difference relative to $\langle C(\kappa^{\rm true})\rangle$, with the gray region representing the 1σ spread of $C(\kappa^{\rm true})$. The full-sky power spectrum is shown in blue. $\overline{C(\kappa^{\rm pix})}$, generated by a forward model that accounts for smoothing but not aliasing, is shown in red. When both effects are included in the forward model (with aliasing parameter $r = 1.6$), $\overline{C(\kappa^{\rm pix})}$ is depicted in purple. Error bars indicate the standard error of the mean. Neglecting either the smoothing or aliasing effects, or both, in the forward model results in an order-of-unity bias on C. 88
3.5 Illustration of the mathematical operations implied by the map-making process (§ 3.4). Plot (1) shows a continuous field. During map-making, the field is first smoothed with a pixel kernel (2). The smoothed field is then sampled at the pixel centers, indicated by white dots in (3). The values at these pixel centers constitute the pixelized map seen in (4). Both the smoothing and sampling processes introduce significant statistical artifacts into the pixelized maps, which must be accounted for in the forward model. 90

3.6 The absolute bias of A_i measured by applying the Gaussian model to Gaussian mocks. Each color corresponds to a Gaussian model with a different aliasing parameter r. Each data point represents the average inferred A_i over 80 independent HSC-like experiments, with the error bars denoting the standard error of the mean. The experiment is set with an 80% margin (see § 3.6.3 and Fig. 3.8). The gray region indicates the 2% systematics control target. 92
3.7 The Kaiser-Squire reconstruction error of a (Gaussian) shear field assuming periodic boundary condition. We define the reconstruction error as $\epsilon = (\mathbb{K}_1\kappa - \gamma_1)/\Delta_\gamma$, where $\mathbb{K}_1$ is the Kaiser-Squire operator that transforms κ to γ_1 assuming a nonperiodic boundary at spatial infinity. Δ_γ is the standard deviation of the shear field. We observe the most error along the boundaries. 93
3.8 The absolute bias of A_i measured by applying the Gaussian model to Gaussian mocks. Each color corresponds to a Gaussian model with a different margin size (as a fraction of the length of the short side of the patch). Each data point represents the average inferred A_i over 80 independent HSC-like experiments, with the error bars denoting the standard error of the mean. The experiment is set with $r = 1.6$ (see § 3.6.2 and Fig. 3.6). The purple points corresponds to the same purple points in Fig. 3.6. The gray region indicates the 2% systematics control target. . . 95
3.9 The impact of the Gaussian transfer map scale-cut, ℓ^G, on the recovered C is presented in terms of the fractional difference relative to the input C. The error bars represent the standard error of the mean. The vertical line denotes the Nyquist frequency, $\ell_{\rm N}$, of $\gamma^{\rm obs}$. Setting $\ell^G = r\ell_{\rm N}$ (red) recovers the input C with the highest accuracy. The observed residual additive bias is related to the limited sky coverage of the patch. 97
3.10 The impact of number density-induced shape noise on C is presented in terms of the fractional difference relative to $\langle C(\kappa^{\rm true})\rangle$ where the number density-induced shape noise is removed (achieved by creating maps using ρ_i and $\gamma^2_{\rm rms}$ at ten times the fiducial value). The high density case is represented in blue, and its 1σ spread is shown in gray (as in Fig. 3.4). $\overline{C(\kappa^{\rm true})}$ calculated with fiducial values for ρ_i and $\gamma_{\rm rms}$ is shown in red. The error bars denote the standard error of the mean. 99

3.11 The joint posterior distribution of A_i inferred from a $40\,\mathrm{deg}^2$ Takahashi patch using the Gaussian (red) and the log-normal (blue) models. The log-normal model employs an informative prior on a_i that is jointly sampled (not shown here) with A_i. Compared to the Gaussian model, the log-normal model exhibits larger uncertainties and a stronger correlation between the A_i's. The contours show the 1σ and 2σ level sets. 102
3.12 The top panel displays the results of the Gaussian model applied to Gaussian, log-normal, and Takahashi mocks. Within this panel, the first row indicates the absolute bias of A_i, averaged over 80 independent HSC-like experiments. The gray area denotes the systematic error control target of 2%. The second row presents the P-P plot for A_i (with the mean bias subtracted), where the x-axis denotes the expected probability (P_{exp}) that a credible interval contains the true A_i, and the y-axis denotes the observed probability (P_{obs}). The vertical dashed line indicates $P_{\mathrm{obs}} = 68\%$, and the colored horizontal lines show where the P-P curves intersect with this line. The Gaussian model demonstrates no absolute bias irrespective of model misspecification but exhibits overconfidence in the error bars when applied to realistic data. The bottom panel illustrates the absolute bias and the P-P plot for A_i using the log-normal model, applied to log-normal and Takahashi mocks. For the Takahashi mocks, three scenarios are tested with the log-normal model concerning the shift parameters: 1) fixed at values measured from noiseless true mocks (fixed a_i), 2) allowed to vary within a narrow and informative range, and 3) allowed to vary within a broad and uninformative range. The log-normal model generally yields accurate uncertainties, yet the absolute bias is highly sensitive to the choice of a_i. 103
3.13 Standardized one-point PDF of $\tilde{\kappa}(\ell)$ for Gaussian, log-normal, and Takahashi mocks at three distinct ℓ values. Across all mocks and ℓ values, the distribution of $\tilde{\kappa}(\ell)$ remains consistently Gaussian, which elucidates the Gaussian model's ability to infer cosmological parameters without absolute bias, notwithstanding model misspecification. 104
3.14 The correlation between $C(\ell_1)/\langle C(\ell_1)\rangle$ and $C(\ell_2)/\langle C(\ell_2)\rangle$ for Gaussian (blue), log-normal (red), and Takahashi (gray) mocks. In this example, we have chosen $\ell_1 = 3340$ and $\ell_2 = 3370$, and C to be the autopower spectrum of the first redshift bin. The contours show the 0.3 and 0.8σ level sets. The Gaussian model assumes no mode correlations. The log-normal model assumes more mode correlation than the realistic mocks. . 107

3.15 Normalized variance of the power spectrum for Gaussian (blue), log-normal (red), and Takahashi (gray) mocks, with C representing the auto-power spectrum of the first redshift bin. Notably, the Gaussian mocks exhibit smaller cosmic variance relative to the Takahashi mocks, whereas the log-normal mocks demonstrate greater cosmic variance in comparison. 108
3.16 The joint posterior distribution of A_i and a_i, as inferred from a Takahashi patch using the log-normal model. We compare an informative prior (red) and an uninformative prior (blue) on a_i. For simplicity, only the first and the last redshift bins are shown. The dashed lines across A_i represent $A_i = 1$, and the dashed lines across a_i represent $a_i^{\rm fit}$—the best-fit values from noiseless data. Three important conclusions are 1) A_i and a_i are positively correlated, 2) the log-normal model prefers a_i much larger than $a_i^{\rm fit}$, and 3) the informative prior truncates the distribution of a_i. Therefore, when the prior is relaxed, A_i shifts to higher values. The contours show the 1σ and 2σ level sets. 110
3.17 Field-level reconstruction for the first redshift bin of a single Takahashi patch using the log-normal (with informative prior on a_i) and the Gaussian models. The first row, from left to right, are the observed noisy shear map $\gamma^{\rm obs}$, the true convergence map $\kappa^{\rm true}$, the mean posterior map $\overline{\kappa^{\rm pix}}$ (log-normal model), the standard deviation map σ (log-normal model), and the S/N map (log-normal model). In the second row, the leftmost plot compares the histograms of $\kappa^{\rm true}$ with those of $\overline{\kappa^{\rm pix}}$ for both Gaussian and log-normal models. The three inset plots zoom in on $1.5\,\deg \times 1.5\,\deg$ region on the patch, contrasting the log-normal results (left) with the Gaussian results (right). 113
3.18 Power spectra is presented in terms of the fractional difference relative to $\overline{C(\kappa^{\rm pix})}$, where $\kappa^{\rm pix}$ are map samples inferred from a single Takahashi patch using the log-normal model with an informative prior on a_i. The power spectrum distribution of the sampled maps, $C(\kappa^{\rm pix})$, is depicted in orange, while the distribution of the mock fields, $C(\kappa^{\rm true})$, is shown in blue to indicate cosmic variance. The power spectrum of the true noiseless field being inferred is shown in black. In this context, C specifies the autopower spectrum of the first redshift bin. The posterior samples are in agreement with $\kappa^{\rm true}$ at the two-point level within the bounds of uncertainty. 114

3.19 Power spectra of the first redshift bin in terms of the fractional difference relative to the adopted theoretical C for the Takahashi mocks (green). The mean of C, as measured on Takahashi mocks, is depicted in blue. The theoretical C, corrected only for shell thickness and resolution, is shown in orange. This correction is found to be insufficient for meeting our accuracy requirements. Abbreviations used: ST for shell-thickness correction, R for resolution correction, and TF for transfer function correction. 117

4.1 We apply HRT to the problem of point mass lensing as described in Section 4.7.1, using a $256 \times 256 \times 512$ box with a mesh resolution $l_{\rm 3D} = 1.66, {\rm Mpc}/h$. We show $\boldsymbol{\alpha}_{\rm HRT}$ in panel (a). For a given $l_{\rm 3D}$, the point mass angular resolution limit is $\theta_{\rm lim} = 4, l_{\rm 3D}/\chi_{\rm l}$ (see Fig. 4.2). We mask light rays within this threshold where the potential field of the point mass lens is not well resolved. The black circle indicates the Einstein radius $\theta_{\rm E}$. Panel (b) illustrates the relative error between $\boldsymbol{\alpha}_{\rm HRT}$ and $\boldsymbol{\alpha}_{\rm theory}$. HRT achieves accuracy within $< 0.5\%$ for both configurations above the resolution limit. 137

4.2 Relative error of $\boldsymbol{\alpha}_{\rm HRT}$ versus $|\theta_0|$ (in unit of $\theta_{\rm lim}$) for two different 3D mesh resolutions and two different lens masses. The low and high mass cases are shown in solid and dashed lines, respectively, with their masses represented in terms of their Einstein radiuses $\theta_{\rm E}$. We observe a universal relationship between the accuracy and $|\boldsymbol{\theta}_0|/\theta_{\rm lim}$. HRT maintains sub-percent accuracy beyond $\theta_{\rm lim}$, demonstrating that mesh resolution is the primary determinant of accuracy. 140

4.3 Convergence (κ) maps generated using the HRT algorithm on a cosmological simulation. The left panel shows the κ map derived from a density field evolved using the standard PM gravity solver. The right panel illustrates the κ map derived from the same initial cosmological conditions but evolved using the PM algorithm enhanced with the spatial optimization (SO) gravity solver, that sharpens the convergence peaks compared to the regular PM. 141

4.4 The top panel shows the distribution of the convergence power spectra obtained by ray tracing through cosmological simulations using HRT. The bottom panel shows the relative error compared to theory (black, solid) for the same cosmology and $n(z)$. The κTNG results are shown in black, dashed lines for comparison. Simulations where gravity is solved using the standard PM method are depicted in blue, while those using the PM enhanced with the spatial optimization (SO) force-sharpening method are shown in orange. The HRT and the κTNG C_ℓ's are similarly suppressed at low ℓ's because of the finite simulation box size and the lack of large-scale modes. The HRT results align with κTNG at large scales but are suppressed at smaller scales since PM methods cannot resolve interactions at or below the mesh resolution. Enhancing PM with SO mitigates this issue and extends C_ℓ^{HRT} to higher ℓ's. For the PM+SO result (where the redshift distribution is a delta function at the same redshift, $z = 1.034$, as the κTNG lensing map), C_ℓ^{HRT} agrees with $C_\ell^{\kappa\text{TNG}}$ to within 10% for $\ell < 800$ and 15% for $\ell < 1200$, where $\ell = 1200$ is marked by the vertical dotted line. 142
4.5 Matter power spectrum at $z = 0$ with the regular PM solver (orange) and with PM + spatial optimization (SO, blue). They are compared to the prediction of the Halofit model in red and the linear theory in green. The shaded regions represent the standard error of the mean estimated with 50 independent simulations. The simulation has 512^3 particles with a 0.4 Mpc/h particle spacing. SO corrects for the small-scale interactions between dark matter particles that are ignored by the regular PM, boosting power at high-k. 145
4.6 Relative error of $\boldsymbol{\alpha}_{\text{HRT}}$ versus θ_{lim} as a function of the number of HRT integration steps used for a point mass lensing task. This figure is to be compared to the blue line in Fig. 4.2 which has the same experimental setup with 45 integration steps. 156
4.7 The effect of ray mesh hyper-parameters: the accuracy parameter ι and the padding size p_{min}. The two panels show $\boldsymbol{\alpha}_{\text{HRT}}$'s residual errors on the image plane for the point mass test. The left panel uses ($\iota = 1,\ p_{\text{min}} = 4$) and the right panel uses ($\iota = 1,\ p_{\text{min}} = 256$). They are compared to panel (b) of Fig. 4.1 which uses ($\iota = 0.5,\ p_{\text{min}} = 256$). 158

4.8 Relative error of the SO-enhanced C_ℓ as a function of the number of HRT integration steps (left) and of the HRT ray mesh accuracy parameter ι (right). The shaded regions show the standard error of the mean over 25 independent simulations. Similar to Fig. 4.4, the black solid line shows the theory prediction, the dashed line shows the κTNG simulation, and the vertical dotted line shows the $\ell = 1200$ scale cut. 159

5.1 *Right:* redshift distribution of the four tomographic bins of the DES Y3 wak lensing sample considered here. *Left:* Power spectra of the 4 noiseless, full-sky convergence maps produced from the FLAMINGO DMO simulation, using the DES Y3 redshift distributions, compared to predictions from EuclidEmu [2]. 174

5.2 Summary statistics of the FLAMINGO κ maps (§ 5.4). Rows correspond to different statistics, and columns to redshift bins (modeled after DES Y3 data). The y-axis shows the FLAMINGO statistics relative to the statistics of the DMO simulation of the same initial condition. On the x-axis, spatial scales decreases to the right. ST and WPH are ordered by the wavelet size (j) as a proxy for their characteristic scale. In FLAMINGO, baryon feedback physics suppresses all statistics on all scales, and is the strongest in low redshift bins. 176

5.3 Summary statistics of the baryonified (colored) and the FLAMINGO fiducial variant's (black, dashed) convergence map, focusing on the first redshift bin. For the baryonified maps, a single BCM parameter is varied at a time. Rows correspond to different BCM parameters, and columns to different statistics. The y-axis shows the statistics of both the baryonified and FLAMINGO maps, relative to those of the DMO simulation with the same initial condition. On the x-axis, spatial scale decreases to the right. ST and WPH statistics are ordered by wavelet size (j), used here as a proxy for characteristic scale. The BCM flexibly generates a range of baryonic signatures in the statistics, clustering around the FLAMINGO results. The variation of other parameters in Tab. 5.1 is shown in Fig. 5.8 in App. 5.9.2. 183

5.4 The emulator mean relative error, grouped by statistics and by redshift bins. The mean absolute (relative) error (MAE) is shown in text. The error is evaluated by averaging across the test set across the baryon parameter space. The emulator achieves sub-percent accuracy for all statistics besides the low redshift moments. The higher-order moment statistics are intrinsically nosier since it involves multiplications of random variables. We remind the reader that the emulator is used to efficiently explore the BCM parameter space when comparing to FLAMINGO statistics. However, the final assessment of how well the BCM reproduces a given FLAMINGO statistic is based on running the full baryonification pipeline at the emulator-derived best-fit parameters. 186
5.5 The residuals error of the BCM when fitted to the FLAMINGO fiducial variant. Rows correspond to different statistics, and columns to redshift bins. The y-axis shows the ratio between the statistics of the best-fit baryonified map to that of the fiducial variant. On the x-axis, spatial scales decreases to the right. ST and WPH are ordered by the wavelet size (j) as a proxy for their characteristic scale. Colors indicate the number of free parameters in the BCM fit. The shaded region marks $\pm 2\%$ for visual reference. For all n-parameter models, the BCM flexibly reproduces the baryonic features in the FLAMINGO fiducial variant simulation to percent level. The prediction is robust even with only 3 degrees of freedoms. Differences in accuracy between the models stem from how we estimate the best-fit parameter, as discussed in § 5.7.2. The residuals for the strongest (fgas-8σ) and weakest (fgas$+2\sigma$) feedback variants are shown in Fig. 5.9 and Fig. 5.10 in App. 5.9.3. 188
5.6 The inferred baryon parameter $\hat{\mathbf{b}}$ from different FLAMINGO feedback variants. Markers represent inference using different degrees of freedom. The color map labels the feedback variants and inherits from Fig. 5.2. In the legend, the feedback strength approximately increases from top to bottom, with the strongest (weakest) feedback model suppression being the fgas-8σ (fgas$+2\sigma$) model. For each feedback variant, the clustering of the points demonstrate the BCM is robust against removing many nuisance parameter dependencies. 194

5.7 The effect of smoothing scale on the two kinds of wavelet coefficients, ST and WPH. Given a statistics $\mathcal{S}$, we define $\mathcal{S}_\ell$ as the statistics measured on κ maps that are low-pass filtered (with a step function) at different ℓ. The statistics measured on the smoothed map is called $\mathcal{S}_\infty$. We test this idea on $\kappa_{\mathbf{dmo}}$, and plot the relative suppression $\mathcal{S}_\ell/\mathcal{S}_\infty$ as a function of smoothing scale ℓ. The line and the error bar represent the mean and the standard error of $\mathcal{S}$ measured on different patches across the full sky. We further color code the lines using the characteristic scale of the wavelet (higher j equals larger wavelets). If $\mathcal{S}_\ell/\mathcal{S}_\infty \ll 1$, then $\mathcal{S}$ contains information beyond ℓ. The $j = 0$ WPH coefficients (shown in red) are suppressed more than 50% at $\ell = 2000$ (shaded region), and hence removed from subsequent analyses. 199
5.8 Same as Fig. 5.3, but for the remaining BCM parameters not shown in the main text. Colored curves show the summary statistics of baryonified convergence maps with one parameter varied at a time, while black dashed curves show the corresponding statistics from the FLAMINGO simulations. 201
5.9 Similar to Fig. 5.5 but with BCM fitted to the strongest feedback variant in the FLAMINGO suite. The BCM error is still at the percent level. The elevated residual for low redshift m_4 is potentially related to both the baryon modelling error and/or m_4's sensitivity to noise. 207
5.10 Similar to Fig. 5.5 but with BCM fitted to the weakest feedback variant in the FLAMINGO suite. The BCM error is still at the percent level. . . 208
5.11 The sensitivity of the baryonic effect to cosmology for the first redshift bin. The y-axis shows the fractional change between the baryonic suppression of summary statistics at the LS8 cosmology compared to the suppression at the fiducial cosmology, for two different subgrid models: the fiducial models (LS8 and L1_m9) and the strong AGN feedback variants (LS8_fgas-8σ and fgas-8σ). Across both feedback variants, the amount of baryonic suppression changes only at the sub-percent level, indicating weak dependence on σ_8. 209

6.1 The h and uncertainty σ maps for **supernovae** (left), **fundamental plane galaxies** (middle) and **CMB** (right). All maps are calibrated to the Planck cosmology using fiducial analysis settings. Color bars mark the 5th, 50th and 95th percentiles. $f_{\rm sky} = 0.41, 0.12, 0.52$. 212

6.2 Pixel value histograms of $\delta_{\rm SN}$, $\delta_{\rm FP}$, and $\delta_{\rm CMB}$. Solid lines show fiducial configurations; dashed lines show systematic variants. The inset shows $\delta_{\rm CMB}$ histograms with error bars indicating 1σ cosmological uncertainty. The distributions are zero-centered and approximately Gaussian. $\delta_{\rm SN}$ has the smallest scatter and $\delta_{\rm CMB}$ the largest. 213
6.3 Weighted C_ℓ (Eq. 6.13) of the Hubble residual maps $\delta_{\rm SN}$, $\delta_{\rm FP}$, and $\delta_{\rm CMB}$. The weighted auto-C_ℓ is scaled by $f_{\rm sky}$ so that the area under the curve approximates the variance of the full-sky δ field. Solid lines represent the data, while shaded regions represent the 1σ intervals of isotropic noise realizations. For the fiducial model (black), $\delta_{\rm SN}$ and $\delta_{\rm CMB}$ are consistent with isotropy, whereas the FP auto-C_ℓ is not, with a significant peak around $\ell = 6$. The SN and FP systematic variants (blue) do not remove peculiar velocities and hence exhibit anisotropic contamination from the large-scale structures. 214
6.4 Upper limits on Hubble constant anisotropy as a function of smoothing scale. Solid lines show the mean noise level from isotropic realizations, while dashed lines indicate 99% confidence upper limits for the SNe, CMB, and the combined maps. 223

7.1 An example of the pointing and advection operators P and W, which map the static CMB signal and the frozen atmosphere screen at $t = 0$ to the TOD, respectively. Rows index TOD samples (t, p) ($t = 0, 1, 2, 3$) and columns index CMB pixels (for P, left panel) or atmosphere pixels at $t = 0$ (for W, right panel); nonzero entries indicate which pixel contributes to each TOD sample. Note that, since the atmospheric screen is moving, the atmosphere map we need to model (7 pixels) extends beyond the spatial domain of the CMB pixels that we observe (4 pixels). 237
7.2 Fourier-space decomposition of the one-dimensional idealized model's TOD. The horizontal axis is spatial frequency k (cycles/pixel) and the vertical axis is temporal frequency ω (cycles/sample); color indicates squared Fourier amplitude (arbitrary units). From left to right, the panels show the atmospheric contribution, the static CMB contribution, and the inverse-covariance filtered TOD $\tilde{N}^{-1}\mathbf{d}$. The drifting atmosphere is confined to the dispersion relation $\omega = -vk$ while the static CMB lies on $\omega = 0$. Applying the optimal filter $\tilde{N}^{-1}$ suppresses power along the atmospheric line but leaves most of the CMB signal invariant. 240

7.3 Uncertainty scaling and eigenmodes of the posterior covariance in the 1D model (top 20 most uncertain modes). The left panel shows the normalized posterior variances, i.e. $1/(\lambda_k \sigma_n^2)$, versus atmosphere-to-noise amplitude ratio. The right panel shows the corresponding eigenmodes, with the same color scale, ordered from most uncertain (bottom) to most certain (top). 243
7.4 Three-dimensional Fourier power decomposition of the idealized 2D TOD example, where the wind is along the $+x$ direction. Axes are spatial frequencies (k_x, k_y) (cycles/pixel) and temporal frequency ω (cycles/sample), with color indicating squared Fourier amplitude (arbitrary units). From left to right, the first panel shows the atmospheric contribution, which lies on the plane $\omega = -\mathbf{k} \cdot \mathbf{v}$ for a frozen screen advected at velocity $\mathbf{v}$ in the $+x$ direction (indicated by the translucent red plane). The middle panel shows the static sky contribution on $\omega = 0$. The right panel shows the inverse-covariance filtered TOD $\tilde{N}^{-1}\mathbf{d}$. The wall projections show the Fourier power integrated over the omitted dimension. Filtering suppresses power near the atmospheric manifold, leaving power concentrated near $\omega = 0$ except along the degeneracy line $\mathbf{k} \cdot \mathbf{v} = 0$. 244
7.5 Representative high-variance eigenmodes of the 2D posterior covariance for a single scan with a fixed wind in the positive x direction. The horizontal axes are pixel coordinates (x, y). Each surface shows one eigenmode of the posterior covariance (ordered from below, from largest to smaller posterior variance), plotted as a height field over (x, y) with color indicating the signed eigenmode amplitude. The least constrained modes are those whose Fourier representation overlaps the single-scan degeneracy $\mathbf{k} \cdot \mathbf{v} = 0$, which for wind along $+x$ corresponds to stripe-like modes with $k_x \simeq 0$ (variation primarily in y). 246
7.6 Idealized reconstruction comparison for two scans with different wind directions, shown in the $\omega = 0$ plane in (k_x, k_y) space (cycles/pixel) with color indicating squared Fourier amplitude (arbitrary units). From left to right, the panels show the true CMB contribution, the reconstruction from a scan with wind along $+x$, the reconstruction from a scan with wind along $+y$, and the optimally combined reconstruction. Each single scan loses modes near its degeneracy line $\mathbf{k} \cdot \mathbf{v} = 0$ (e.g. $k_x = 0$ for wind along $+x$), while combining scans with different wind directions restores these modes up to the monopole. 247

LIST OF TABLES

1.1 The relationship between the topics of field-level inference and the structure of the present thesis. 5

2.1 The redshift parameters. 48
2.2 The cosmological parameter constraints given by the CMB-ISW-galaxy model. Here $A^{\mathcal{P}}$ and $A^{\mathcal{I}}$ are the amplitudes of the primordial and the ISW power spectra and b_i are the tomographic galaxy biases. 54

4.1 Definitions of mesh and ray variables. The white rows define the resolution of meshes or rays, while the light gray ones define their corresponding position vectors. The top two rows are for the 3D particle mesh, on which the gravitational forces are computed. The middle two rows specify the 2D configuration of rays. And the bottom two rows describe the 2D ray mesh on which we interpolate and transfer the transverse gravitational forces, from the 3D particle mesh and to the 2D ray positions. 134

5.1 The input parameters to the profiles of the BCM. For each parameter, we list its prior range and a description of the parameter's physical meaning. See Anbajagane *et al.* [3] for the specific equations that these parameters govern. The prior ranges are similar to those in [3, 4] and are used in § 5.6.1 to construct samples of the baryon-corrected maps and to train the BCM emulator. 170
5.2 Summary of statistics used in the analysis, restricted to scales with $\ell \lesssim 2000$. 182
5.3 The best-fit BCM parameters obtained from the joint fitting of all HOS to different FLAMINGO variants. These parameter combinations yield HOS that agree with FLAMINGO measurements to 2%. Other BCM parameters are fixed to $\nu_{M_c} = 0$, $\nu_{\theta_{\rm ej}} = 0$, $\gamma = 2.60$, $\log_{10}\theta_{co} = -1.00$, $\log_{10}\eta_\delta = -0.28$, $\delta = 14.0$, and $\mu_\beta = 1.60$. The uncertainty is the 16% and 84% credible interval of our fitting method. See Tab. 5.4 in App. 5.9.4 for the best-fit values for all n-parameter models. 191
5.4 The best-fit BCM parameters obtained from the joint fitting of all HOS to different FLAMINGO variants. These parameter combinations yield HOS that agree with FLAMINGO measurements to 2%. The integers following the feedback variant names are the number of free parameters. The fixed parameters are highlighted in red. The uncertainty is the 16% and 84% credible interval. 204

6.1 Results of the isotropy χ^2 test using SNe, CMB, and FP tracers. The main result uses only the SNe and the CMB data (row 1), and supports the null hypothesis of isotropy. Lower rows show the test statistics when varying a single tracer's systematic model or combining with the FP data. 220

7.1 The TOD forward model for a two-pixel, two-timestep experiment. We denote the static CMB values as c_1, c_2 and the atmospheric screen at $t = 0$ have a single degree of freedom a_0. **Left:** The atmosphere drifts to the right by one pixel per timestep. **Right:** The atmosphere is static relative to the sky. 230

ACKNOWLEDGMENTS

Science is written by human beings. When we stand before the unknown, I suspect it is no more purely objective than any other mode of reading the world—no more detached than interpretation itself. Yet in this, there is something profoundly beautiful. For we perceive always through the senses, and the senses are inseparable from the inner life that animates them. Memory, fear, intuition, doubt, wonder, and hope. If our surroundings are gentle, sustaining, and humane, then I believe the same humanity will find its way—quietly but surely—into the questions we ask, and into the equations, papers, and sciences we bring into being. Thus some theories and experiments come through as calm and rigorous, others exuberant, daring, and still others impulsive, rash. Once articulated, these ideas take on a kind of permanence and are subjected to the long judgment of history.

Ideas, like living things, bear the imprint of their origins. If a thought is seeded with gentleness and care, then surely it will grow into something gentle and caring in return. The universe is full of life, and it is through life, and by life, that we come to see and understand the universe at all. In this sense, science is not merely the discovery of something etched immutably into the mechanical code of nature.

This is the conviction with which I have tried to carry the past three and a half years of research. I hope that the contents of this thesis speak of humanity as much as they speak of logic; that in attempting to read the book of nature, I have translated—through my own sensibilities and moments of excitement—something that may endure careful examination and reward thoughtful reflection. As such, this work stands as both my conviction and insistence: that science remains among

the most profoundly human of endeavors, and that its terrible responsibilities and wondrous promises are burdens we must learn to carry together.

I wish first to express my deepest gratitude to my advisor, Scott Dodelson. Long before I could imagine myself here, I remember reading (struggling through) your book *Modern Cosmology* in a cafe during the pandemic's lockdown. It planted a seed of curiosity about the universe that has continued to grow to this day. That I should later have the unimagined fortune to study under your guidance is still deeply humbling. From Carnegie Mellon to the University of Chicago, I thank you for your unconditional trust and your rare, deeply empathetic support throughout my studies. There is a quality—best captured, perhaps, by the Chinese image of 天马行空, a winged and harnessless horse roaming freely through the stars—that characterizes the intellectual freedom you offered me: a freedom without neglect, an openness without indifference. You gave me the courage to ask, and to attempt to answer, the hardest questions we can pose to the universe. I am certain that this courage will continue to guide me as I try to build the world I hope to see. Thank you, above all, for your contagious kindness—to me and to the communities around you—and for your patience with my growth, both as a scientist and as a person.

I would also like to thank the other members of my thesis committee—John Carlstrom, Chihway Chang, and Peter Littlewood—for their generosity of insight, time, and intellectual care.

I have been equally fortunate to learn from many remarkable collaborators. I thank my undergraduate advisor, Michael Troxel, for introducing me to cosmology and for cultivating a sense of scientific curiosity and independence that has remained

a constant guide. I am deeply grateful to Rachel Mandelbaum for exemplifying the highest standards of scientific rigor—standards that continually motivated me to pursue not only what science can discover, but how it ought to be done. From you, I learned that uncompromised integrity, curiosity, and honesty can and should coexist with the softest forms of encouragement and care; that precision and accuracy demand not coldness, but great empathy and courage.

I would also like to thank the many others with whom I had the fortune to work and write, in no particular order: Xiangchong Li, Marco Gatti, Dhayaa Anbajagane, Yuuki Omori, Chunhao To, Yin Li, Dan Scolnic, Tianqing Zhang, Sunao Sugiyama, Hannah Fronenberg, Biwei Dai, Nianyi Chen, Yihao Zhou, Lei Hu, Bonny Wang, Antoine Bourdin, and my office mates both in Pittsburgh and in Chicago.

I wish to thank my family, and especially my mother, for twenty-five years of unwavering and unconditional trust and support. I hope that what is enclosed in this thesis belongs to you as much as it belongs to me, and that it may make you proud.

Finally, I thank my friends—Jennifer, Jupiter, Joyce, Yue, Lor, Medy, Ventali, Andrew, Michael, Coco, Yuncong, Oscar, Alice, and Aixuan. The past years have not been easy. It was your companionship, both in person and at a distance, that carried me through this journey.

To all, thank you for showing me a world of kindness and warmth. With that, let me close with this thought:

Because the world treated me with gentleness,

I grew into gentleness myself.

I now offer that gentleness back,

hoping it may become gentler still.

CHAPTER 1

INTRODUCTION

Cosmology seeks to understand the composition, origin, and evolution of the Universe by interpreting the faint traces of information carried to us by light [5–7]. Our observations begin as time-ordered detector measurements or as raw images taken by telescopes. These data are then processed through map-making or catalog-building pipelines, yielding maps of the sky at different electromagnetic wavelengths or catalogs of detected objects such as stars and galaxies [8–13]. These observables are manifestations of the underlying cosmological matter fields whose full structure we never observe directly. From these partial, and often noisy and incomplete glimpses, we attempt to reconstruct both the physical laws that govern cosmic evolution and the initial conditions from which present-day structures emerged. This task is an inverse problem of exceptional scale – the phenomena we wish to infer are weak, nonlinear, and even chaotic, while the data we receive are heterogeneous, noisy, and incomplete. The challenge for both us and cosmology is therefore to connect these two realms in a principled and quantitatively controlled manner.

For much of its history, modern cosmological inference has relied on a classical paradigm that separates observation, compression, and interpretation. Telescopes first produce raw data that is converted into maps or catalogs. These intermediate products are then condensed into a set of summary statistics – means and variances, or more precisely, correlation functions or power spectra [10, 14, 15]. These condensed summary statistics are useful because they are deterministic functions of the cosmological parameters, and can be computed deterministically through first

principles. Therefore, in the final stage of a cosmological analysis, we compare the observed statistics with analytical results through likelihood-based inference to estimate the underlying cosmological parameters. This procedure has yielded many of the central results of modern cosmology.

This framework is effective for several reasons. First, the compression from maps to summary statistics is computationally fast, reducing millions of pixels or billions of objects to an $\mathcal{O}(10^2)$–dimensional data vector. Second, the mapping from cosmological parameters and experimental nuisance parameters to these statistics is analytically tractable, at least on large scales, allowing forward modeling to be performed reliably and repeatedly. Third, because this forward model is simple, it is straightforward to invert: likelihood evaluations are fast and the inverse problem can be solved with standard optimization or sampling techniques [16, 17]. These features explain why the traditional workflow has endured and supported the quantitative success of modern cosmology.

In symbolic form, the classical workflow may be summarized by a pair of two-way relations connecting maps, summary statistics, and cosmological parameters.

$$\text{maps/catalogs} \underset{\text{inverse inference}}{\overset{\text{compression}}{\rightleftharpoons}} \text{summary statistics} \underset{\text{inverse inference}}{\overset{\text{easy to model}}{\rightleftharpoons}} \text{cosmological parameters.} \tag{1.1}$$

The likelihood function adopted in this setting takes the schematic form

$$\log \mathcal{L} \propto (S_{\text{data}} - S_{\text{th}}(\theta, \nu))^T \, C^{-1} \, (S_{\text{data}} - S_{\text{th}}(\theta, \nu)) + \log \text{prior}(\theta, \nu), \tag{1.2}$$

where $S_{\rm data}$ denotes the measured summary statistics data vector, $S_{\rm th}(\theta, \nu)$ their theoretical predictions for cosmological parameters θ and nuisance parameters ν, and C is the associated covariance matrix.

Yet the unidirectional compression from maps to summary statistics is not lossless. Non-Gaussian structures in the maps are often projected onto a small set of low-order moments. Indeed, as illustrated explicitly in Chapter 3, two fields that differ strikingly in their morphology—one containing coherent non-linear features and another lacking them—can nevertheless map to identical summary-statistic [18, 19]. This illustrates that compression can discard cosmologically relevant information, even on the scales where structure formation is most informative.

These limitations motivate a reformulation of cosmological inference in which compression is omitted altogether. In the field-level paradigm, we replace summary statistics with the data themselves [18, 20–23]. In the most ambitious scheme of field-level cosmological inference, one directly compares each pixel (or time-ordered datum) directly to the output of a forward model that deterministically evolves the initial conditions of the Universe, incorporates the relevant physical processes, and constructs the observables appropriate to each experiment. Cosmological parameters, nuisance parameters, and the initial conditions themselves are treated as joint unknowns. The corresponding likelihood takes the schematic form

$$\log \mathcal{L} \propto (d - f(\theta, \phi_{\rm init}, \nu))^T \, N^{-1} \, (d - f(\theta, \phi_{\rm init}, \nu)) + \log \text{prior}(\theta, \nu, \phi_{\rm init}), \quad (1.3)$$

where d is the data vector, f is the forward model, and $\phi_{\rm init}$ denotes the approximately Gaussian initial density at the early epochs of the Universe. In this for-

mulation, map-making, cosmological inference, and multiprobe combination are performed simultaneously. The field-level formulation dispenses with summary statistics altogether and relates the data directly to the parameters through the forward model.

$$\text{maps/catalogs/time-ordered data} \underset{\text{forward model}}{\overset{\text{inverse}}{\underset{\text{inference}}{\rightleftharpoons}}} \text{cosmological parameters.} \tag{1.4}$$

The absence of an intermediate summary-statistic stage in this diagram is intentional: all information available at the map level is, in principle, preserved and leveraged.

Three immediate consequences follow. First, field-level inference retains the full hierarchy of N-point information, avoiding the irreversibility of compression [18, 22, 24]. Second, by placing the initial conditions on the same footing as the cosmological parameters, the method enables simultaneous reconstruction of the present-day Universe and of the primordial density field from which it evolved [20]. Furthermore, because the evolution from the initial density field to the late-time structure is governed by deterministic gravitational dynamics, knowing the initial condition implies, in principle, knowledge of the entire four-dimensional trajectory of structure formation. Thus, the inference of initial conditions is simultaneously the inference of cosmic history. Third, the framework naturally incorporates multiple probes, since each observable is generated from the same underlying fields through different physical response operators.

The example above jointly infers the cosmological parameters and the initial conditions of the density fields using field-level observations. However, this field-level modeling and inference paradigm has much broader implications for the entire

Analytical	**Non-Gaussian fields**		**Map making**
	Approximate models	More realism	
Statistics of FL solutions (Chs. 2 [24], 7)	Phenomenological weak lensing (Ch. 3 [18])	Differentiable simulations (Ch. 4 [26])	Isotropy of the cosmic expansion (Ch. 6 [25])
Multiprobe analysis (Ch. 2 [24])	FL systematics (Ch. 3 [18])	Baryon corrections (Ch. 5 [19])	

Table 1.1: The relationship between the topics of field-level inference and the structure of the present thesis.

problem space where we have basic understanding of the physical system, and we wish to extract latent model parameters from noisy, incompletely, and multi-modal spatial temporal data.

To understand what can be recovered in such an approach, it is necessary to examine the mathematical structure of the field-level posterior itself. This is most easily studied in regimes where the likelihood is exactly known and the forward model is analytically tractable. This is shown in the first column of Tab. 1.1. Chapters 2 and 7 study precisely such regimes: the joint reconstruction of the primordial CMB, the integrated Sachs–Wolfe signal, and galaxy density fields on large scales; and the removal of atmospheric contamination in ground-based CMB observations. In both cases, the analysis illuminates how biases arise at the field level, how posterior sampling corrects these biases, and how multiprobe information flows through the inference [24].

Beyond these analytically controlled examples lies the more general situation encountered in weak gravitational lensing, where the underlying matter field is non-

Gaussian and the forward model is approximate. Here the question is whether field-level inference can be extended to realistic data, and what sources of systematic error arise when the true fields deviate from the assumed priors. This is the second focus of the thesis, and is shown in the second column of Tab. 1.1. Chapter 3 addresses these questions by examining the behavior of approximate forward models on lensing maps, identifying new sources of field-level systematics, and quantifying their effect on cosmological parameters. The analysis shows that both the structure of the prior and the details of map construction influence the accuracy of the inference, and it establishes a regime in which these effects can be controlled at a level suitable for ongoing surveys [18].

The phenomenological models arise out of computational comprise. The exact physical forward model for cosmological structure formation is actually known in principle – on large spatial scales, the governing dynamics are precisely those of general relativistic gravity. If the initial density field were known, the subsequent evolution of matter would be fully determined. However, solving this forward model through numerical simulation is extremely expensive. If one imagines a setting with effectively unbounded computational resources, the optimal cosmological analysis would evolve the initial conditions forward under gravitational dynamics, construct the observables for each experiment, and compare them directly to the data, thereby performing field-level inference with an exact physical model. Realistically, two obstacles must be overcome: the forward model must be fast enough to be evaluated repeatedly, and it must be differentiable with respect to both the initial conditions and the cosmological parameters, since realistic inference in these high-dimensional

spaces requires gradients for sampling and optimization. [17, 27] Chapter 4 (the first row of the third column in Tab. 1.1) addresses these challenges by developing a differentiable N-body framework that evolves matter and photon trajectories jointly, without the thin-lens approximation, using a Hamiltonian formulation designed for stability, accuracy, and compatibility with automatic differentiation [26].

Even with such a forward model in hand, uncertainties remain on small and intermediate spatial scales where baryonic processes reshape the matter distribution. Hydrodynamical simulations produce divergent predictions for these scales, reflecting the inherent uncertainty in modeling star formation, feedback, and gas dynamics. Chapter 5 (the second row of the third column in Tab. 1.1) approaches this problem by constructing a flexible semi-analytic correction model capable of reproducing a broad range of baryonic effects across both two-point and non-Gaussian statistics. This approach provides a computationally efficient means of marginalizing baryonic uncertainties in field-level, summary-statistic, and simulation-based analyses alike [19].

Finally, the field-level perspective enables direct tests of the symmetries underlying the standard cosmological model. If the expansion of the Universe is isotropic, then directional variations in the measured Hubble rate should be consistent with noise and sample variance, and uncorrelated across epochs. Chapter 6 (the rightmost column in Tab. 1.1) constructs full-sky maps of expansion rate variation using Type Ia supernovae, fundamental plane galaxies, and the CMB, and examines both their individual anisotropies and their cross-correlations. This analysis provides a map-level test of the isotropy of the cosmic expansion from redshift zero to the surface of

last scattering [25].

These are the directions I have enjoyed exploring during the past three and a half years. This thesis synthesizes those developments and the methods they motivate, with the aim of providing a coherent account of how cosmological information may be reconstructed at the level of fields.

CHAPTER 2

FIELD-LEVEL MULTIPROBE ANALYSIS OF THE CMB, INTEGRATED SACHS-WOLFE EFFECT, AND THE GALAXY DENSITY MAPS

2.1 Abstract

Extracting information from cosmic surveys is often done in a two-step process, construction of maps and then summary statistics such as two-point functions. We use simulations to demonstrate the advantages of a general Bayesian framework that consistently combines different cosmological experiments on the field level, and reconstructs both the maps and cosmological parameters. We apply our method to jointly reconstruct the primordial CMB, the integrated Sachs-Wolfe effect, and six tomographic galaxy density maps on the full sky on large scales along with several cosmological parameters. While the traditional maximum a posterior estimator has both two-point level and field-level bias, the new approach yields unbiased cosmological constraints and improves the signal-to-noise ratio of the maps.

2.2 Introduction

The large-scale structure (LSS) of the Universe is defined by the full 3-dimensional matter density field $\delta(\mathbf{x}, t)$. Although it is difficult to determine $\delta(\mathbf{x}, t)$ directly, we extract information about it indirectly in two general ways: (i) light from distant sources (including the cosmic microwave background) is impacted by over and un-

derdense regions; and (ii) gravitationally bound objects such as galaxies and clusters often *trace* the matter density. Examples of the first class of information include the late-time integrated Sachs-Wolfe (ISW) effect caused by decaying gravitational potentials in the dark-energy era and the deflection of photons due to gravitational lensing. The second class includes galaxy clustering and cluster counts. One important objective of modern cosmology is to develop statistical methods to combine this information in the most efficient and consistent manner, in order to reconstruct $\delta(\mathbf{x}, t)$ and constrain models of its origin and evolution.

In the past decades, independent experiments have made extraordinary advances in charting these individual tracers. For example, on the cosmic microwave background (CMB) front, several generations of anisotropy and polarization measurements have led to recent results; the Planck Collaboration has mapped the temperature and polarization anisotropy of the early Universe and used its lensing statistics to study the integrated gravitational potential along the line of sight [28, 29]. The Atacama Cosmology Telescope (ACT) and the South Pole Telescope (SPT) have made similar achievements with smaller footprints but higher resolutions [30–33]. Stage-III wide-field photometric surveys such as the Dark Energy Survey (DES), the KioDegree Survey (KiDS) and the Hyper Suprime-Cam (HSC) have observed millions of galaxies on a significant fraction of the sky and used galaxy positions and shape statistics to probe the low-redshift matter distributions [10, 34, 35]. The recipe for analyzing most of this data involves first converting the data into 2-dimensional maps (e.g., for CMB surveys) and catalogs (e.g., for galaxy surveys); computing the correlation functions (or the power spectra) of these fields, and then comparing

these observed correlation statistics to a cosmological model in a Bayesian likelihood analysis to yield cosmological parameter constraints. In almost all of these cases, the fiducial cosmological model, ΛCDM, fits the data well.

In addition to these results from single probes, there has been an increased effort to maximize information by combining probes. An example of this is the recent DES result combining its data of galaxy positions and galaxy shapes with the projected gravitational potential measured by SPT and Planck [36]. In this example, roughly the same recipe is followed: DES made maps of the galaxy density in five tomographic bins and the shear in four bins; these were combined with maps of the projected gravitational potential from SPT and Planck. Given these three sets of maps, there are six sets of two-point functions (galaxy clustering, galaxy-galaxy lensing; cosmic shear, cosmic shear $\times$ CMB lensing, galaxy density $\times$ CMB lensing, and the CMB lensing auto-correlation function). This set of six two-point functions forms the data vector, which is then used to constrain parameters. The main goal of this effort is to extract from all this low-redshift (much lower than the decoupling of the CMB) data a measurement of the amount of clustering at late times. This is often quantified with S_8, which the DES+SPT analysis determined to be $S_8 = 0.792 \pm 0.012$, lower than the Planck measurement, $S_8 = 0.832 \pm 0.013$. The discrepancy does not meet strict statistical standards but it has spawned much interest and it is reminiscent of the Hubble tension that is driven by different measurements of the zeroth order expansion rate of the universe.

Taking stock, the fiducial cosmological model fits most of the data, but there are alluring hints that it is flawed, and one of the most intriguing ways of stress-testing

the model is to measure how the clustering of matter evolves over the course of time. To date, this has been done predominantly by: (i) map-making, (ii) compression to two-point functions, and (iii) parameter constraints.

Research into field-level analysis offers an opportunity to change the way that we extract data from surveys, in the process offering an alluring opportunity for a powerful suite of tests of ΛCDM. The basic idea of field-level analysis is to combine all three steps above into one. Early examples of this idea [37–42] focused on the CMB. In that example, the time-ordered data can be converted into a map at the same time that the power spectrum is determined. The parameters to fit for the data, therefore, are the values of the temperature in all the pixels in the map plus a handful of cosmological parameters that determine the power spectrum. Eriksen *et al.* [40] extended the idea to allow for multiple maps to be constructed: e.g., maps of foregrounds in addition to the CMB. This basic technology has been incorporated into the most recent results from Planck [43]. Groups are now applying the technology to galaxy surveys [44–49].

One way to understand the advantage of the field-level approach is to return to the DES+SPT analysis: first CMB lensing maps were made using the traditional quadratic estimator [50] and then they were used to construct two-point functions. However, the data in DES itself could in principle help improve the fidelity of the CMB lensing maps: after all, the deflection of the CMB photons is due (at least in part) to the very structure that DES measures. Combining this information would clearly create a better CMB lensing map. Using that improved map with DES maps though would be a form of double counting, so it makes sense to do everything at

once: create all the maps and estimate all the power spectrum simultaneously. In the particular example of CMB lensing, the problem is not trivial but Millea *et al.* [22, 51, 52] have made significant progress simultaneously measuring the lensing field, the primordial CMB, and several parameters that determine the relevant power spectra.

Here we use simulated data sets on large scales to (i) develop the machinery that can handle real data; (ii) explore some of the basics of field-level analyses; and (iii) provide an example of how the field-level analyses can be used to stress-test ΛCDM. Our example is related to the work in Eriksen *et al.* [40], except that we attempt to separate the late-time Integrated Sachs-Wolfe (ISW) signal from the primordial CMB anisotropies. Hang *et al.* [53] constrained the ISW and lensing amplitudes using the two-point correlation between the DESI Legacy Survey and Planck temperature and lensing maps, where as we are interested in the full posterior distribution of both the parameters, two-point functions, and the maps.

We begin in §II by explaining some of the details; then in §III, we analyze simulated CMB data assuming that it consists only of noise and CMB anisotropies. We recover some of the known problems of the maximum posterior solution (the Wiener filter) and show that these can be mitigated by instead using samples of the full posterior. Then, in §IV, we introduce the ISW component and try to separate that from the primordial anisotropies. The degeneracies make this problematic at the map level, but the sampler produces an unbiased power spectrum for each. This is crucial, as the cosmological parameters themselves are embedded in the spectrum so if the spectrum is unbiased, then the parameters will be as well. Specifically, we

introduce two free amplitudes of each component that multiply the fiducial spectra and show that the field-level analysis that simultaneously solves for the map values and the parameters produces unbiased estimates of the parameters.

The ensuing constraints on the amplitude of the ISW spectrum are not very restrictive, so in §V, we explore the possibility of adding in other tracers, the galaxy density in several tomographic bins. This adds to the number of free parameters in the field-level analysis but we show that it produces a higher fidelity ISW map and a fairly tight constraint on the amplitude of the ISW spectrum. This leads to the prospect of stress testing ΛCDM by introducing amplitudes in front of all spectra (CMB lensing; galaxy density; cosmic shear) in addition to the standard cosmological parameters: a measurement in which any one of these amplitudes is determined to deviate from unity will disprove ΛCDM by demonstrating that structure does not grow in time as predicted by the model.

In short, our goal in this paper is to explain (to some, much of this will not be new) what to expect when carrying out a field-level analysis; demonstrate how well it does on simulated data with increasing numbers of components and probes, and point the way to a simple but powerful way to stress test the fiducial cosmological model. We share our conclusions and thoughts about the next steps in §VI.

2.3 Theoretical framework

2.3.1 Field-level multiprobe analysis

The general problem of a field-level multiprobe inference is summarized in Fig. 2.1. The data is an observation, or a set of observations, on the sky. For concreteness, we will focus on the synergy between CMB experiments and photometric galaxy surveys, but the argument generalizes to any combinations of probes.

The data is assumed to consist of a set of signals and noise:

$$d = \sum_{\alpha} s^{\alpha} + n. \tag{2.1}$$

Our model assumes that the signals s^{α} in the data are drawn from a Gaussian distribution with mean zero and covariance matrix, $\mathbb{C}^{\alpha\beta}(\theta)$, where θ represents cosmological and nuisance parameters. The noise is also drawn from a Gaussian distribution with mean zero and known covariance matrix $\mathbb{C}^{n}$.

The likelihood for obtaining the data given the cosmological parameters and the signals is

$$-2\ln\mathcal{L} = \left[d - \sum_{\alpha} s^{\alpha}\right], [\mathbb{C}^{n}]^{-1}, \left[d - \sum_{\beta} s^{\beta}\right] + \dots \tag{2.2}$$

where the additional terms are irrelevant, and the products on the right involve all pixels. That is, in the case of a single survey with $N_{\rm pix}$ pixels, d is a set of the values in all the pixels, and $\mathbb{C}^{n}$ is a $N_{\rm pix} \times N_{\rm pix}$ matrix. If only one signal contributes, then s also has $N_{\rm pix}$ values; if more signals are assumed, then the total number of parameters in all the s^{α} will be $N_{\rm pix} \times N_{\rm signal}$. When data from multiple surveys

are used, d will be a concatenated version of all the individual data sets and different signals can contribute to different data sets.

Using the Bayes theorem, we can invoke the prior on all the signals and the parameters. Since we are confining our analysis to large scales throughout, the prior on all signals is Gaussian, and the posterior is

$$\begin{aligned} -2\ln p = &-2\ln\mathcal{L} + \sum_{\alpha\beta} s^{\alpha}\left(\mathbb{C}^{-1}(\theta)\right)^{\alpha\beta} s^{\beta} \\ &+ \ln\det\mathbb{C}(\theta) - 2\ \ln \mathrm{prior}(\theta) \end{aligned} \tag{2.3}$$

where irrelevant terms have been dropped. The parameters in this posterior are θ (which determines $\mathbb{C}$) and the map(s) s^{α}. For example, in the case of a single survey, if there is one signal contributing and there are 5 cosmological parameters, then the number of parameters we use to fit the $N_{\rm pix}$ data points is $N_{\rm pix}+5$. There are often cases where there are two or more signals contributing. For example, below we model the CMB as consisting of the signal from the last scattering surface plus the contribution from the late-time ISW effect. In that case, there will be $2N_{\rm pix}+5$ free parameters.

As described in §II.C, we will draw samples from this posterior. The accumulated samples of both the maps and the cosmological and nuisance parameters are fully consistent in the Bayesian sense. More precisely, the distribution of the values of map pixels s^{α} will provide a set of posterior samples of the signals, and the distribution of the parameters will constrain the relevant models of interest. These distributions

will be consistent with one another, so that for example in a sample with a large $\mathbb{C}^{\alpha\alpha}$, the signal s^α everywhere is likely to have a larger dispersion.

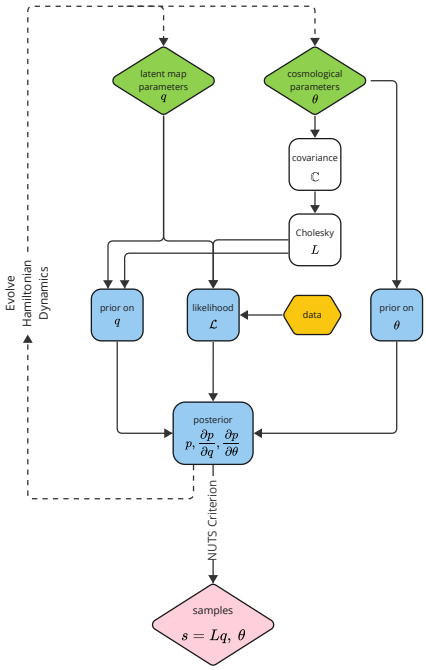

Figure 2.1: The flow chart of a general field-level multiprobe analysis that accumulates samples of both the cosmological parameters and the tracer maps (values of each signal in each pixel) as discussed in §2.3.1 and §2.3.3. We start from the prior distribution of the cosmological and latent map parameters. We then use the realized cosmological parameters to construct the covariance of the tracers, which in turn transforms the latent map parameters into physical tracer maps. The covariance, tracer maps, and observed data are then combined into the likelihood and – after multiplying by the priors – the posterior functions. If the above calculations are all programmatically differentiable, we can calculate the derivatives of the posterior function easily, and use HMC No-U-Turn Sampler (HMC-NUTS) to efficiently sample from the very high dimensional posterior space. In this diagram, diamonds denote sampled parameters, squares denote model-relevant functions, and the hexagon is the (fixed) observed data vector. The pink diamond represent samples that represents the posterior space.

2.3.2 Pixels

Above we glossed over the details of the map. Here, we review the basics of pixels in terms of the coefficients of spherical harmonics and explain why we choose to work with this basis.

Consider a map on the curved sky $s(\mathbf{n})$, where $\mathbf{n}$ is a 3-dimensional unit vector. Analogous to Fourier transformations in Euclidean spaces, we can study this field in the frequency (or harmonics) space via forward and inverse spherical harmonics transform (SHT),

$$s(\mathbf{n}) = \sum_{lm} s_{lm} Y_{lm}(\mathbf{n}) \tag{2.4}$$

$$s_{lm} = \int \frac{d\Omega}{4\pi} s(\mathbf{n}) Y^*_{lm}(\mathbf{n}) \tag{2.5}$$

where $Y_{lm}(\mathbf{n})$'s are the set of orthonormal spherical harmonics. We adopt the HEALPix pixelization strategy (where the angular resolution is specified by a single parameter `NSIDE`), and use the discretized SHT as implemented by the `healpy` library [54, 55]. As usual in cosmological analyses, we drop the monopole and dipole modes ($l = 0, 1$).

In general, if the field s is statistically homogeneous and isotropic, it is more advantageous to study s's correlation structure in harmonic space. In real space, the correlation function between two line-of-sight directions is given by

$$w_{\mathbf{n},\mathbf{n}'} = w(|\mathbf{n} - \mathbf{n}'|) = \langle s(\mathbf{n}) s(\mathbf{n}') \rangle \tag{2.6}$$

where we see that the correlation function has dense off-diagonals. For a discretized map with NSIDE resolution, the size of w scales as $\texttt{NSIDE}^4$, which quickly becomes impossible to handle (for example, an $\texttt{NSIDE} = 256$ map has an angular resolution of $27''$ and 8×10^5 pixels; the full pixel-pixel covariance matrix totals 5 terabytes).

However, s's power spectrum (s's correlation function in harmonic space), $\mathbb{C}$, defined by

$$\langle s^{\alpha}_{lm} s^{*,\beta}_{l'm'} \rangle = \delta_{ll'} \delta_{mm'} \mathbb{C}^{\alpha\beta}_{l} \tag{2.7}$$

is diagonal in this basis and depends only on the multiple moment l and the different sets of signals assumed. Therefore, the amount of memory needed to manipulate $\mathbb{C}$ is linear in NSIDE. One important caveat to this simplicity is that the field must be homogeneous and isotropic, and these assumptions fail in the presence of instrumental noise patterns, partial sky coverage, and masking.

2.3.3 Methodology

Here we present the details of our implementation of Fig. 2.1. The fundamental idea behind all MC sampling techniques is: start from the current sample; find the next point in the parameter space and generate a probabilistic proposal to make it a sample (both operations may involve repeated evaluations of the posterior density). How to find the next point and what proposal to make are algorithm-specific, however, they in general satisfy the principle of detailed balance such that, in the limit of large sample size, the samples approximate the posterior distribution.

The efficiency of MC sampling rests on the suitability of the MC algorithm for the specific inference context and the effective computation of the posterior distribution.

The efficiency of MC sampling rests on the suitability of the MC algorithm for the specific inference context and the effective computation of the posterior distribution.

For the first point, since we are inferring both the map pixels and the cosmological parameters, the dimensionality of the posterior space will be quite large. For example, for our final analysis in §V which includes 8 tracer maps at `NSIDE` $= 32$, the total dimensionality of the posterior space is 73704. This is too large for traditional Monte Carlo techniques (such as Metropolis–Hastings) to operate efficiently. Intuitively, this is because as dimensionality increases, the ratio between the neighboring volume pointing towards and away from a particular point in the parameter space (e.g., the mode of the distribution) decays exponentially. Thus, the Random Walk Metropolis algorithm becomes overwhelmingly likely to propose samples outside the typical set, where the target density and hence the acceptance probability vanishes [17].

Hamiltonian Monte Carlo (HMC) solves this efficiency problem in high dimensional spaces [17, 56, 57]. In the HMC framework, we augment the parameter space with a conjugate momentum space and use the gradient of the log posterior surface to guide us to sample only near the bulk of the probabilistic mass [17, 56]. In order to avoid traditional HMC's sensitivity to hyper-parameters such as the integration steps, we further employ the No-U-Turn Sampler (NUTS) variation of the HMC, first proposed by Hoffman and Gelman [57].

This leads us to the second point on computational efficiency. HMC samplers require repeated evaluations of the posterior function and its gradient. Since we want to develop a multiprobe field-level framework that easily extends to different

observables and cosmological models, we do not want to hard-code the derivatives in advance. Instead, we choose to make the framework pragmatically differentiable through the JAX auto-differentiation library in python, which interfaces smoothly with the `numpyro` implementation of the NUTS [58, 59].

Turning to the specific problem of the posterior function computation. We start at the top of Fig. 2.1 and break this calculation into several parts:

- We start with a proposal for the cosmological parameters θ and the latent map parameters. The latent map parameters are a set of uncorrelated standard Gaussian variables q which we will later transform into the signal maps.
- Calculate $\mathbb{C}$ given the cosmological parameters
- Transform the latent map parameters using the Cholesky decomposition ($\mathbb{C} = LL^t$) of the covariance matrix: $s = Lr$. (The prior of the maps becomes $s^2/(2\mathbb{C}) = q^2/(2\mathbb{I})$).
- Combine the maps s with the data to calculate the likelihood (forward modeling)
- Use the likelihood and the prior to calculate the posterior and its derivative with respect to the parameters
- If the NUT criterion is satisfied, accept this as a valid sample
- Use the leapfrog method to generate another sample

In practice, through the JAX auto-differentiation library, this framework provides information on both the posterior and its gradient.

The treatment of the covariance function deserves more discussion. In this paper, we will keep the shape of the spectra fixed and allow for free amplitudes $(A^\alpha)^2$. We assume the amplitudes have fiducial values equal to one and have a uniform prior distribution. We then construct the full covariance matrix $\mathbb{C}^{\alpha\beta}$, which consists of the auto- and cross-spectra of each signal. In principle, since $\mathbb{C}$ encodes the covariance between all the pixels for all the tracer maps, its dimensionality is very high. For a single HEALPix map at the resolution of `NSIDE` (or a limiting resolution of $l_{\rm max} = 3\texttt{NSIDE} - 1$), there are

$$\sum_{l=2}^{l_{\rm max}} (2l+1) = l_{\rm max}^2 + 2l_{\rm max} - 3 \tag{2.8}$$

degrees of freedom ignoring monopole and dipole modes. Again, for our final analysis in §V, which includes 8 tracers, the size of $\mathbb{C}$ is on the order of $8^2 \times l_{\rm max}^4$. The efficient computation of this covariance matrix is one of the limiting factors in the feasibility of field-level analysis. However, in the limit of full sky and when all the fields are homogeneous and isotropic, the sub-block of $\mathbb{C}$ for each tracer is diagonal in the a_{lm} basis. Thus, we can bring $\mathbb{C}$ into block diagonal forms, with $l_{\rm max} - 2$ unique $\mathbb{C}_l$ sub-blocks on the diagonal. Each $\mathbb{C}_l$ sub-block has size 8×8, describing the correlation between the 8 tracers at mode l. Looking ahead, we will be considering the primordial CMB signal (modulated by $A^{\mathcal{P}}$); the late-time ISW effect (modulated by $A^{\mathcal{I}}$); and the galaxy density in 6 tomographic bins (modulated by b_i's). The first

of these is uncorrelated with the rest, so the ensuing 8×8 sub-block matrix will be

$$\mathbb{C}_l(b_1, ..., b_6, A^{\mathcal{I}}, A^{\mathcal{P}}) = \begin{pmatrix} b_1^2\mathbb{C}_l^{\mathcal{G}_1,\mathcal{G}_1} & b_1 b_2\mathbb{C}_l^{\mathcal{G}_1,\mathcal{G}_2} & ... & b_1 A^{\mathcal{I}}\mathbb{C}_l^{\mathcal{G}_1,\mathcal{I}} & 0 \\ b_1 b_2\mathbb{C}_l^{\mathcal{G}_2,\mathcal{G}_1} & b_2^2\mathbb{C}_l^{\mathcal{G}_2,\mathcal{G}_2} & ... & b_2 A^{\mathcal{I}}\mathbb{C}_l^{\mathcal{G}_2,\mathcal{I}} & 0 \\ ... & ... & ... & ... & \\ b_1 b_6\mathbb{C}_l^{\mathcal{G}_6,\mathcal{G}_1} & b_2 b_6\mathbb{C}_l^{\mathcal{G}_6,\mathcal{G}_2} & ... & b_6 A^{\mathcal{I}}\mathbb{C}_l^{\mathcal{G}_6,\mathcal{I}} & 0 \\ b_1 A^{\mathcal{I}}\mathbb{C}_l^{\mathcal{I},\mathcal{G}_1} & b_6 A^{\mathcal{I}}\mathbb{C}_l^{\mathcal{I},\mathcal{G}_2} & ... & (A^{\mathcal{I}})^2\mathbb{C}_l^{\mathcal{I},\mathcal{I}} & 0 \\ 0 & 0 & \dots & 0 & (A^{\mathcal{P}})^2\mathbb{C}_l^{\mathcal{P},\mathcal{P}} \end{pmatrix} \tag{2.9}$$

With this computationally efficient representation of the covariance matrix, we can transform the latent map variables into the proper tracer maps through either the sub-blocks' Cholesky representations or their eigen-decomposition. We experimented with both, and found the former to be an order of magnitude faster (see also Loureiro *et al.* [45]). In summary, the algorithm is very fast. For the largest model we considered in §2.6, the analysis was done on an Apple M1 chip running overnight.

2.4 Reconstruction of the primordial CMB map

We start with the simplest possible example, a single simulated CMB all-sky map. Although simple, this model demonstrates the key behaviors of two ways of using the posterior: identifying the free parameters by finding the point at which the posterior is maximum (hereafter, maximum a posteriori or MAP) and generating samples of the posterior (hereafter sampling). We compare the potential biases of

both methods and discuss the implication for the field, two-point, and cosmological parameter constraints. In some ways, the idea of asking whether an estimator is biased is introducing frequentist ideas into a Bayesian discussion. Nonetheless, we think that understanding these biases is an important step towards the ultimate goal of extracting the correct cosmological parameters from the data. The intuition we find here will serve us well in the subsequent more complex cases.

The primordial CMB temperature fluctuations $s^{\mathcal{P}}$ originate from the time of recombination ($z_* \approx 1100$), when the photons decoupled from the photon-electron-proton fluid as the universe cooled below a few percent of the ionization energy of hydrogen. We assume (and all current data is consistent with this assumption, with the tightest constraints coming from the Planck Collaboration *et al.* [60]) that the resulting temperature variation is a homogeneous and isotropic random Gaussian field which is fully characterized by the power spectrum $\mathbb{C}_l^{\mathcal{P}}$. Throughout this paper, we assume a fiducial cosmology of $H_0 = 100h = 67.5\text{km/Mpc/sec}$, $\Omega_b h^2 = 0.0219$, $\Omega_c h^2 = 0.1139$, $A_s = 2 \times 10^{-9}$ and $n_s = 0.965$.

2.4.1 Biases of the optimal estimator

Fixed cosmological parameters

The observed temperature data $d^{\mathcal{T}}(\mathbf{n})$ is the superposition of the primordial field $s^{\mathcal{P}}(\mathbf{n})$ and noise $n^{\mathcal{T}}(\mathbf{n})$.

$$d^{\mathcal{T}}(\mathbf{n}) = s^{\mathcal{P}}(\mathbf{n}) + n^{\mathcal{T}}(\mathbf{n}) \tag{2.10}$$

We call this the CMB model. We ask: Given $d^{\mathcal{T}}$ and perfect knowledge of $s^{\mathcal{P}}$'s and $n^{\mathcal{T}}$'s theoretical power spectra ($\mathbb{C}^{\mathcal{P}}$ and $\mathbb{C}^{n,\mathcal{T}}$ respectively), how well can we reconstruct the primordial field? Additionally, how accurate is the power spectrum of the reconstructed field?

In the Bayesian framework, the posterior probability in Eq. 2.3 reduces to

$$-2\ln p(s^{\mathcal{P}}|d^{\mathcal{T}}) \propto \sum_{lm}\left(\frac{|d^{\mathcal{T}}_{lm}-s^{\mathcal{P}}_{lm}|^2}{\mathbb{C}^{n,\mathcal{T}}_l}+\frac{|s^{\mathcal{P}}_{lm}|^2}{\mathbb{C}^{\mathcal{P}}_l}\right) \tag{2.11}$$

where we drop the determinant terms since the CMB and noise spectra are assumed known and fixed. The MAP solution for $s^{\mathcal{P}}$ is then given by the Wiener filter

$$\hat{s}^{\mathcal{P},\mathbf{MAP}}_{lm} = \frac{\mathbb{C}^{\mathcal{P}}_l}{\mathbb{C}^{\mathcal{P}}_l+\mathbb{C}^{n,\mathcal{T}}_l} d^{\mathcal{T}}_{lm}. \tag{2.12}$$

The mean power spectrum of the MAP estimator

$$\langle\hat{\mathbb{C}}^{\mathcal{P},\mathbf{MAP}}_l\rangle \equiv \langle\frac{1}{2l+1},\sum_m|\hat{s}^{\mathcal{P},\mathbf{MAP}}_{lm}|^2\rangle = \frac{\mathbb{C}^{\mathcal{P}}_l}{\mathbb{C}^{\mathcal{P}}_l+\mathbb{C}^{n,\mathcal{T}}_l}\mathbb{C}^{\mathcal{P}}_l \tag{2.13}$$

is known to be biased [61, 62], an effect more prominent in the low signal-to-noise ratio (SNR) regime.

To implement, we simulate $s^{\mathcal{P}}$ from the fiducial cosmology power spectrum on a HEALPix grid of `NSIDE` = 64 and then inject isotropic white noise $n^{\mathcal{T}}$ with a relatively high variance of $\mathrm{Var}(n^{\mathcal{T}}) = 1000\mu K^2$. Even though this exceeds noise in Planck by several orders of magnitude, we use this value to demonstrate the difficulties of extracting the signal in the presence of appreciable noise. The power

spectra of the truth map (black) and the recovered MAP map (green) are shown in Fig. 2.2. As the amplitude of the noise spectrum (purple) rises on small scales, the MAP spectrum is increasingly suppressed. On the field level, this means that $\hat{s}_{lm}^{\mathcal{P},\mathbf{MAP}}$ is damped for over and underdensities on scales that have small SNR. The estimator does not have an additive bias but does have a multiplicative bias, i.e., $\langle \hat{s}_{lm}^{\mathcal{P},\mathbf{MAP}} / s_{lm}^{\mathcal{P}} \rangle = \mathbb{C}_l^{\mathcal{P}} / (\mathbb{C}_l^{\mathcal{P}} + \mathbb{C}_l^{n,\mathcal{T}})$ for both the real and the imaginary components. This is shown in the top panel of Fig. 2.3.

Varying cosmological parameters

In real cosmological analyses, we are also interested in cosmological parameters (such as the primordial amplitude, spectral index, etc.) that modify the shape and amplitude of the power spectrum. We want to know how the field, two-point, and parameter MAP estimators behave when the spectrum is allowed to change.

For example, consider modulating the fiducial power spectrum $\mathbb{C}^{\mathcal{P}}$ with a scale-invariant amplitude A^2, where A has a flat prior. The new MAP solutions are given by

$$\hat{s}_{lm}^{\mathcal{P},\mathbf{MAP}} = \frac{(\hat{A}^{\mathbf{MAP}})^2 \mathbb{C}_l^{\mathcal{P}}}{(\hat{A}^{\mathbf{MAP}})^2 \mathbb{C}_l^{\mathcal{P}} + \mathbb{C}_l^{n,\mathcal{T}}} d_{lm}^{\mathcal{T}} \tag{2.14}$$

$$\hat{\mathbb{C}}_l^{\mathcal{P},\mathbf{MAP}} = \frac{(\hat{A}^{\mathbf{MAP}})^2 \mathbb{C}_l^{\mathcal{P}}}{(\hat{A}^{\mathbf{MAP}})^2 \mathbb{C}_l^{\mathcal{P}} + \mathbb{C}_l^{n,\mathcal{T}}} (\hat{A}^{\mathbf{MAP}})^2 \mathbb{C}_l^{\mathcal{P}} \tag{2.15}$$

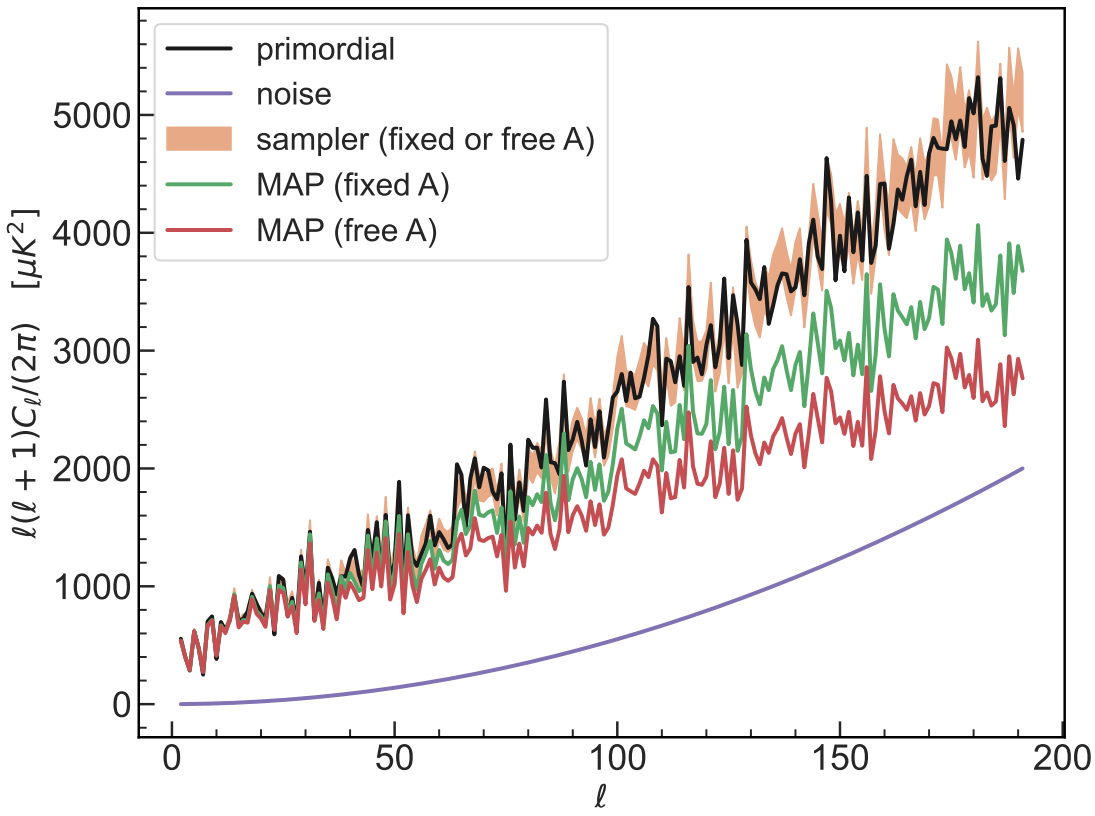

Figure 2.2: A numerical experiment of CMB reconstruction that demonstrates the difference between the MAP estimators and the sampling-based estimators. We generated the truth map from a fiducial $\mathbb{C}^{\mathcal{P}}$ and added isotropic noise. The truth and noise spectra are shown in black and purple. We test two reconstruction models, the first has a fixed $\mathbb{C}^{\mathcal{P}}$ (Eq. 2.11), and the second has a free amplitude A^2 modulating the fiducial $\mathbb{C}^{\mathcal{P}}$. Their MAP solutions are given by Eqs. 2.13 and 2.15, and are shown in green and red respectively. We confirm the analytical solutions match the results from direct numerical optimization. Both two-point and MAP estimators are biased lower than truth and the free amplitude model has a greater bias. We then sampled $s^{\mathcal{P}}$ directly from the posterior space. The spectra of the mean maps for the fixed and free amplitude cases are biased (both overlap the green curve). However, in both cases, the distributions (the orange shaded region represent 1σ credible interval) of the spectra scatter around the truth and are unbiased.

and $\hat{A}^{\mathbf{MAP}}$ satisfies

$$\sum_l (2l+1)\left\{1 - \frac{(\hat{A}^{\mathbf{MAP}})^2 \mathbb{D}_l \mathbb{C}_l^{\mathcal{P}}}{((\hat{A}^{\mathbf{MAP}})^2 \mathbb{C}_l^{\mathcal{P}} + \mathbb{C}_l^{n,\mathcal{T}})^2}\right\} = 0 \tag{2.16}$$

where $\mathbb{D}_l$ is the data power spectrum.

When noise is present, $\hat{A}^{\mathbf{MAP}} < 1$, in this case equal to 0.81 (see Fig. 2.4). Thus, by comparing Eqs. 2.13 and 2.15, we see that the new MAP is biased even lower than the truth. We can apply this model to the same reconstruction experiment as before. The result for the new power spectrum estimator is shown in red in Fig. 2.2. The multiplicative bias in the field-level estimator is shown in the bottom panel of Fig. 2.3.

For power spectra with complicated parameter dependence, we often lack analytical optimal solutions. However, qualitatively speaking, if an increase in the parameter increases the amplitude of the spectrum as in this case, then the parameter will be underestimated by optimal inference, and vice versa.

2.4.2 Sampling the CMB field

Fixed cosmological parameters

Now we seek an unbiased estimator for $\mathbb{C}^{\mathcal{P}}$ that also has a convenient notion of uncertainty. Let us again first fix $A = 1$ and draw a sample of maps $\{s^{\mathcal{P}}_{lm,i}\}_{i=1,\ldots,N}$ directly from the posterior distribution (Eq. 2.11). Using this set of maps, we can

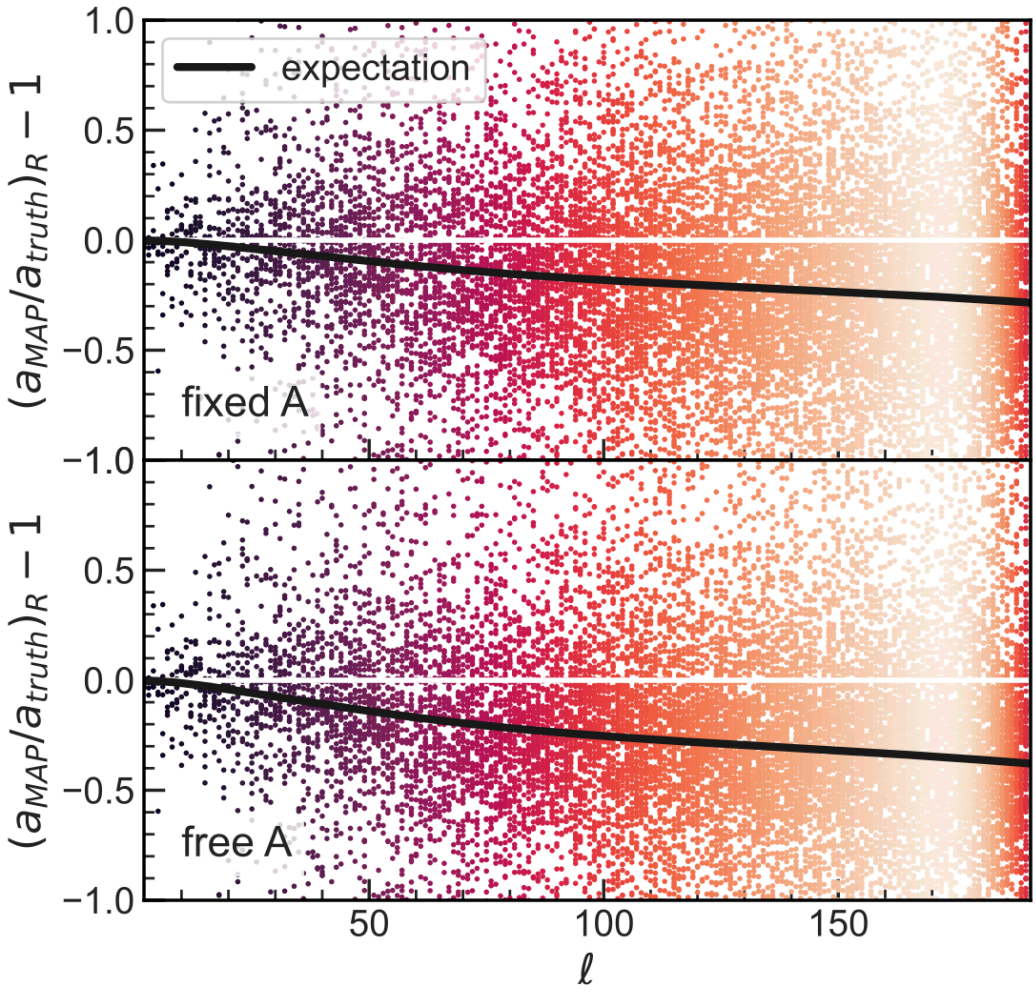

Figure 2.3: The scatter plot of the ratio between the real components of the MAP a_{lm} and that of the truth a_{lm} as a function of l. The cases for fixed and varying amplitude parameters are shown in the upper and lower panels respectively. In each case, the mean of the ratios (black curve) matches the Wiener filter expectations (with the MAP amplitude in the case of varying amplitude). The color map represents point cloud density.

construct an associated set of power-spectra samples

$$\{\mathbb{C}^{\mathcal{P}}_{l,i}\} = \{\sum_m \frac{1}{2l+1}|s^{\mathcal{P}}_{lm,i}|^2\} \tag{2.17}$$

Let us call the ensemble average of $\{s^{\mathcal{P}}_{lm,i}\}$ and $\{\mathbb{C}^{\mathcal{P}}_{l,i}\}$ as $\overline{s^{\mathcal{P}}}$ and $\overline{\mathbb{C}^{\mathcal{P}}}$, respectively. It is crucially important that the power spectrum of $\overline{s^{\mathcal{P}}}$ is *different* from $\overline{\mathbb{C}^{\mathcal{P}}}$.

We claim, in the limit of sufficient sample size N,

1. $\overline{s^{\mathcal{P}}}$ and its spectrum are exactly the field and two-point MAP estimators (Eqs.2.12 and 2.13), and they have the Wiener filter multiplicative bias.

2. On the two-point level, the samples $\{\mathbb{C}^{\mathcal{P}}_{l,i}\}$ give a proper Bayesian credible interval centered around the truth.

3. Further, $\overline{\mathbb{C}^{\mathcal{P}}}$ (and more generally, the mean of any n-point power spectrum samples) is an *unbiased* estimator in the frequentist sense (when we have multiple data realizations).

We prove these claims in Appendix 2.8.2. However, intuitively, how can the power spectrum of the mean map be biased while the sampled power spectrum be unbiased? One way to understand this is to think of each sampled map as $s^{\mathcal{P}} = \hat{s}^{\mathcal{P},\mathbf{MAP}} + s'$. When we compute the power spectrum $\langle|s^{\mathcal{P}}|^2\rangle$, the $\langle|s'|^2\rangle$ term exactly compensates for the deficiency of the MAP spectrum. Alternatively, the $s^{\mathcal{P}}$'s are normally distributed, and for any Gaussian distribution the mean is equal to the maximum, so $\overline{s^{\mathcal{P}}}$ is the MAP solution. However, the power spectrum is not normally distributed; its expected value is an unbiased estimator of truth, and not the biased MAP solution.

We continue with the numerical experiment above. This time, we construct an HMC NUT sampler following the prescription of §2.3.3, using Eq. 2.11 as our posterior distribution. After the chain equilibrates, we draw 3000 $s^{\mathcal{P}}$'s from the posterior space. We confirm that the power spectrum of the mean map exactly follows the MAP solution for the case of fixed parameters (the green curve in Fig. 2.2). We further show the distribution of the sampled power spectra $\{\mathbb{C}_{l,i}^{\mathcal{P}}\}$ in Fig. 2.2 in shaded orange. Indeed, the distribution of the spectra covers the truth power spectrum within uncertainty, and $\overline{\mathbb{C}^{\mathcal{P}}}$ is unbiased.

We note that similar phenomena have been observed in previous studies. For example, in Fig. 8 of [52], the authors find that the distribution of the sampled CMB spectra scatter around truth while the spectrum of the mean map is biased lower at small scales.

The above observations have the following implications. One must debias the sampled maps before using them for cosmological analysis, similar to how we currently correct for MAP maps (e.g., with analytical or Monte Carlo-based corrections). However, if we are only performing analysis on the two-point level, the samples are *unbiased* and their distribution constitutes a convenient measure of uncertainty. In short, by considering the sampled power spectra, we recapture an unbiased estimator of $\mathbb{C}^{\mathcal{P}}$.

Varying cosmological parameters

The exact two-point statistics recovery motivates us to ask whether the sampled cosmological parameters that modify the power spectrum are also unbiased. To

answer this question, we use the HMC NUT sampler from the previous section with an additional A^2 modulating the fiducial spectrum. We assume A has a flat and wide prior on $[0.2, 10]$ and collect 3000 samples after appropriate burn-in.

We find that A is unbiased with its marginal distribution shown in Fig. 2.4. We can also estimate the variance of its distribution, which is predicted by the inverse of the Fisher information

$$\mathcal{F}^{-1} = \sum_l \frac{2(2l+1)}{1 + (\mathbb{C}^{n,\mathcal{T}}/\mathbb{C}^{\mathcal{P}})^2}, \tag{2.18}$$

shown as shaded orange in the same figure.

Since the expectation of A is unbiased, it follows that the posterior samples $\{\mathbb{C}^{\mathcal{P}}_{l,i}\}$ again scatter around truth unbiased. In fact, their distribution overlaps that of the fixed amplitude model almost exactly as shown in Fig. 2.2 in shaded orange. Further, the power spectrum of the mean map is also not the MAP anymore; it is the MAP solution of the model with fixed cosmological parameter (Eq. 2.13), as if A is fixed to 1.

The important takeaway is the following. The MAP map and the mean sampled map are biased both on the field level and on the two-point level. However, the distribution of the sampled power spectra (and cosmological parameters that modulate them) is unbiased.

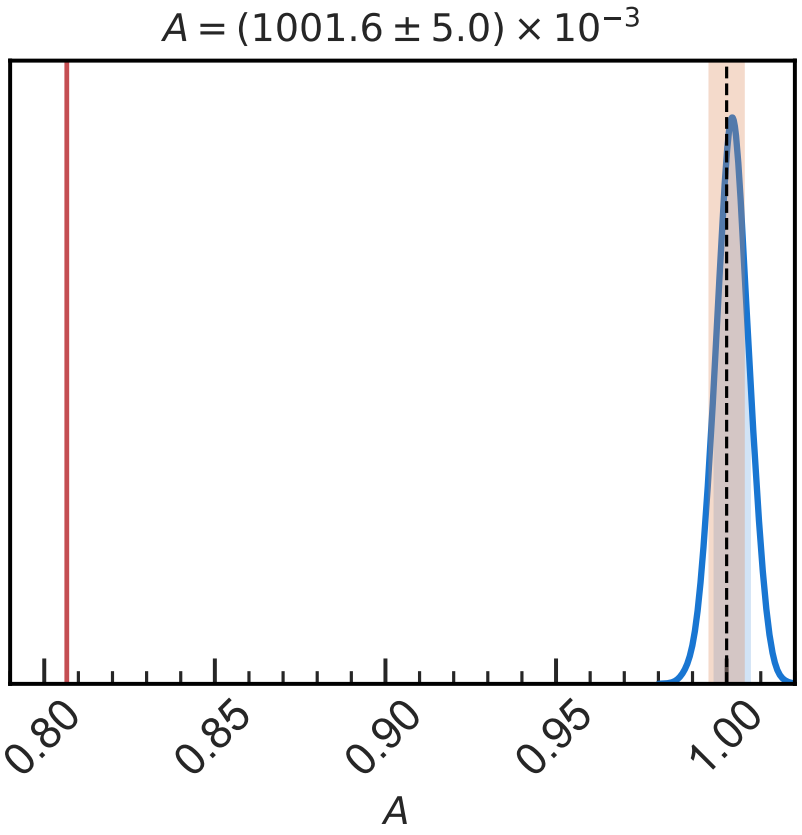

Figure 2.4: The posterior distribution of the amplitude A (blue curve) compared to truth (black dashed line). The shaded orange area represents the Fisher forecast of the 1σ uncertainty centered on truth, and the shaded blue area represents the 1σ uncertainty reported by the sampler. The red line shows the MAP value for A, which is quite far from truth.

2.5 Joint reconstruction of the primordial CMB and the ISW effect

Now, we expand on the CMB model and consider extracting the primordial and the ISW contributions from a single noisy temperature measurement. As we shall see, the MAP estimators for both signals are again biased on the field and the two-point level. The sampled fields have a multiplicative bias but their two-point statistics are unbiased. The new challenge in this case study is the field-level degeneracy between the primordial and the ISW maps, which motivates the multiprobe approach presented in the next section [40]. We discuss the key properties of this degeneracy, which we expect to be quite general when one separates a low SNR map from a measurement based on a likelihood approach.

2.5.1 The ISW effect

In the late universe, the primordial CMB fluctuations are modified by the ISW effect on very large scales [63–67]. A photon is blue-shifted when descending a gravitational potential well and red-shifted when it escapes. When the universe began its accelerated expansion (in the dark energy-dominated era), the large-scale potential wells decayed. As a result, a photon will leave the decaying well (barrier) with more (less) energy than it enters. The observed ISW temperature modification is thus the integrated effect of the decaying potential well along the line of sight, and

its 2-dimensional field is given by [68, 69]

$$A^{\mathcal{I}}(\mathbf{n}) = \int_0^\infty \frac{\partial\Phi(\mathbf{x}[\chi,\mathbf{n}],t(\chi))}{\partial t}\frac{2e^{-\tau(\chi)}}{1+z(\chi)}d\chi \tag{2.19}$$

Here, $\tau(\chi)$ is the optical depth out to distance χ and Φ is the 3-dimensional gravitational potential field, which ultimately depends on the matter overdensities $\delta_m(\mathbf{x}[\chi,\mathbf{n}],t(\chi))$. By rewriting Φ in terms of δ_m, and moving to the Fourier space, we can rewrite Eq. 2.19 as

$$a_{lm}^{\mathcal{I}} = 4\pi i^l \int \frac{d^3k}{(2\pi)^3} I_l^{\mathcal{I}}(k) Y_{lm}^*(\mathbf{k})\delta_m(\mathbf{k},t_0) \tag{2.20}$$

where

$$I_l^{\mathcal{I}}(k) = \int d\chi D(\chi) W^{\mathcal{I}}(k,\chi) j_l(k\chi) \tag{2.21}$$

and the window function is given by

$$W^{\mathcal{I}}(k,\chi) = -\Theta(\chi_* - \chi)\frac{3\Omega_m H_0^2}{k^2}\frac{\partial \ln((1+z)D(z))}{\partial t} \tag{2.22}$$

where $D(z)$ is the growth function normalized to unity at $z = 0$ and we approximate τ as zero through the epoch that the ISW is generated.

The forms of Eqs. 2.20 - 2.22 are not peculiar to the ISW effect - by modifying the window function $W(\chi,k)$ appropriately, the 2-dimensional observable of most tracers can be computed as a line-of-sight integral of δ_m. For general tracers, A and B, of this form, the covariance is

$$\mathbb{C}_l^{A,B} = \frac{2}{\pi}\int_0^\infty k^2 dk P(k) I_l^A(k) I_l^B(k). \tag{2.23}$$

One way of quantifying the correlation between different probes is to compute the scale-dependent correlation coefficients, defined as

$$\rho^{AB} = \frac{\mathbb{C}^{AB}}{\sqrt{\mathbb{C}^{AA}\mathbb{C}^{BB}}}. \tag{2.24}$$

We again assume both fields are statistically homogeneous and isotropic, so the covariance is diagonal in the a_{lm} basis. For example, the primordial CMB and the ISW effect are spatially independent ($\mathbb{C}^{\mathcal{I},\mathcal{P}} = 0$), and their autopower spectra are shown in Fig. 2.5 ($\mathbb{C}^{\mathcal{I},\mathcal{P}} = 0$). The ISW signal is primarily confined to the very large scales ($l < 10$) that enter the horizon during the dark energy-dominated era. The ISW signal is also subdominant to the primordial signal on all scales, making it particularly challenging to reconstruct. We explore the case of nondiagonal covariance in §2.6 in the context of multiprobe joint reconstruction.

2.5.2 Separating the primordial and ISW signals

The observed temperature data $d^{\mathcal{T}}(\mathbf{n})$ is the sum of the primordial field $s^{\mathcal{P}}(\mathbf{n})$, the ISW field $s^{\mathcal{I}}(\mathbf{n})$, and noise $n^{\mathcal{T}}(\mathbf{n})$, so the posterior distribution

$$\begin{aligned}
&\ln p(s^{\mathcal{P}}, s^{\mathcal{I}}, A^{P}, A^{I} | d^{\mathcal{T}}) \propto \\
&\sum_{lm} \frac{1}{2} \left(\log {A^{\mathcal{P}}}^2 \mathbb{C}_l^{\mathcal{P}} + \log {A^{\mathcal{I}}}^2 \mathbb{C}_l^{\mathcal{I}} \right) \\
&+ \sum_{lm} \left(\frac{|d_{lm}^{\mathcal{T}} - s_{lm}^{\mathcal{P}} - s_{lm}^{\mathcal{I}}|^2}{2\mathbb{C}_l^{n,\mathcal{T}}} + \frac{|s_{lm}^{\mathcal{P}}|^2}{2{A^{\mathcal{P}}}^2 \mathbb{C}_l^{\mathcal{P}}} + \frac{|s_{lm}^{\mathcal{I}}|^2}{2{A^{\mathcal{I}}}^2 \mathbb{C}_l^{\mathcal{I}}} \right)
\end{aligned} \tag{2.25}$$

We will refer to this model as the CMB-ISW model.

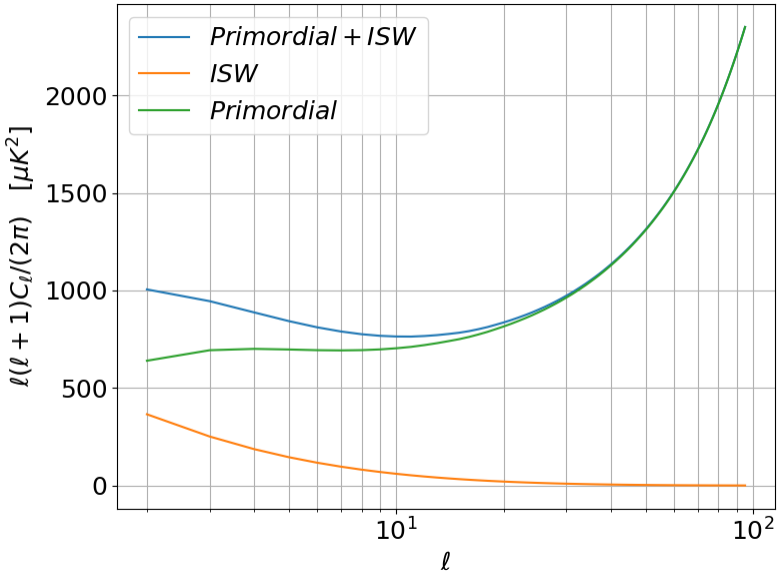

Figure 2.5: The power spectra of the primordial CMB and the ISW effect on large scales (low-ℓ modes).

One crucial difference between the CMB and the CMB-ISW model is that although both are constrained by the same amount of data $d^{\mathcal{T}}$, the dimensionality of the latter's posterior space ($s^{\mathcal{P}}$ and $s^{\mathcal{I}}$) is (ignoring the cosmological parameters) twice than that of the former's (only $s^{\mathcal{P}}$). In other words, if we have N independent modes on the full sky, we are trying to constrain $2N$ parameters with N data points in the CMB-ISW model. Hence, we expect significant degeneracy in the inferred $s^{\mathcal{P}}$ and $s^{\mathcal{I}}$ maps, and we want to explore how this affects the MAP and the sampling-based field-level reconstructions. We emphasize that this problem will be quite common in any field-level analysis where we wish to separate the different physical components of a single observed field.

We will again tackle this problem in two ways, first by constructing the MAP

estimators and then by sampling directly from the posterior distribution.

Fixed cosmological parameters

Let us first fix $A^{\mathcal{P}} = A^{\mathcal{I}} = 1$ and seek the field-level MAP solutions. When we are trying to find the MAP of $s^{\mathcal{P}}$, the effective noise is the sum of the instrumental noise and the ISW temperature fluctuation (and analogously for $s^{\mathcal{I}}$). Thus, invoking the Wiener filter (Eq. 2.12),

$$\hat{s}^{\mathcal{P},\mathbf{MAP}}_{lm} = \frac{\mathbb{C}^{\mathcal{P}}_l}{\mathbb{C}^{\mathcal{P}}_l + \mathbb{C}^{n,\mathcal{T}} + \mathbb{C}^{\mathcal{I}}} d^{\mathcal{T}}_{lm} \tag{2.26}$$

$$\hat{s}^{\mathcal{I},\mathbf{MAP}}_{lm} = \frac{\mathbb{C}^{\mathcal{I}}_l}{\mathbb{C}^{\mathcal{I}}_l + \mathbb{C}^{n,\mathcal{T}} + \mathbb{C}^{\mathcal{P}}} d^{\mathcal{T}}_{lm} \tag{2.27}$$

and similarly for their power spectra.

As in §2.4, we simulate this reconstruction method numerically by generating $s^{\mathcal{P}}$ and $s^{\mathcal{I}}$ with fiducial cosmology a HEALPix grid of `NSIDE` = 32 with $\mathrm{Var}(n^{\mathcal{T}}) = 200\ \mu K^2$. In the top panels of Fig. 2.6, the truth power spectra are shown in black, the MAP power spectra in green, and the effective noise in purple. The primordial MAP spectrum is biased low both on large scales (ISW contamination) and small scales (noise contamination). The ISW MAP spectrum is significantly biased low on all scales due to the same Wiener filter suppression.

On the field level, we again expect (and indeed observe) no additive bias but a multiplicative bias on the $\langle \hat{s}^{\mathbf{MAP}}_{lm} / s_{lm} \rangle$ proportional to the Wiener filter factor (for both the real and the imaginary components). The case for the ISW field is particularly egregious, as shown in the top panel of Fig. 2.7 (note that the y-axis

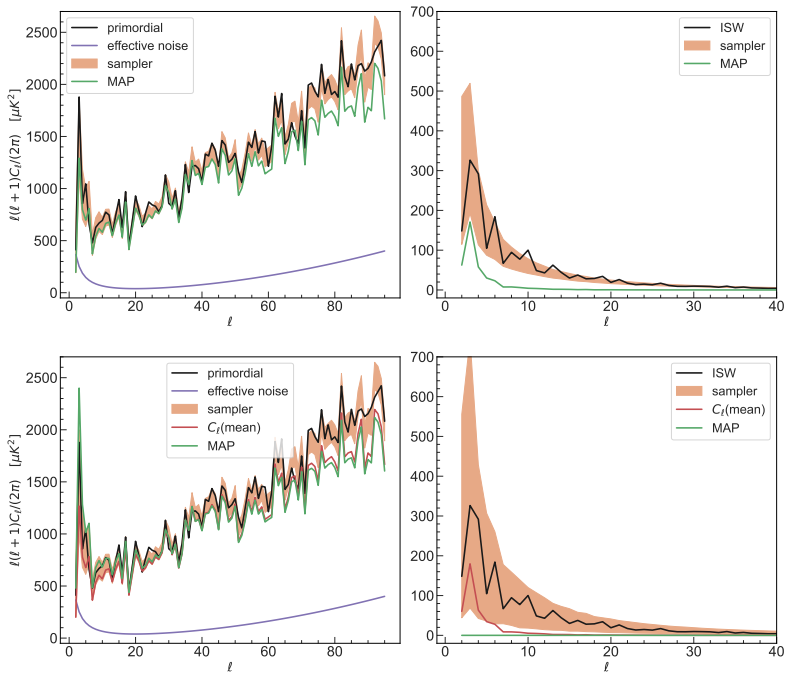

Figure 2.6: The recovered primordial ($s^{\mathcal{P}}$) and ISW ($s^{\mathcal{I}}$) spectra using the various models in §2.5.2. The simulation is drawn from the fiducial cosmology and a noise variance of 200 μK^2. The top and bottom panels show the case of fixed and free amplitude respectively, and the left and right panels show the primordial and the ISW results, respectively. For each case, the truth spectrum is shown in black, the MAP result is shown in green, and the 1σ credible interval of the sampled spectra is shown in shaded orange. The effective noise of the primordial maps is $\mathbb{C}^{\mathcal{I}} + \mathbb{C}^{n,\mathcal{T}}$ shown in purple. The effective noise of the ISW map ($\mathbb{C}^{\mathcal{P}} + \mathbb{C}^{n,\mathcal{T}}$) is above the ISW spectra in the right panels on all scales and hence not shown in the plot. Notice that in the case of free amplitude, the MAP solution is different from the $\mathbb{C}_l$ of the sampled maps (red).

does not even contain the unbiased case).

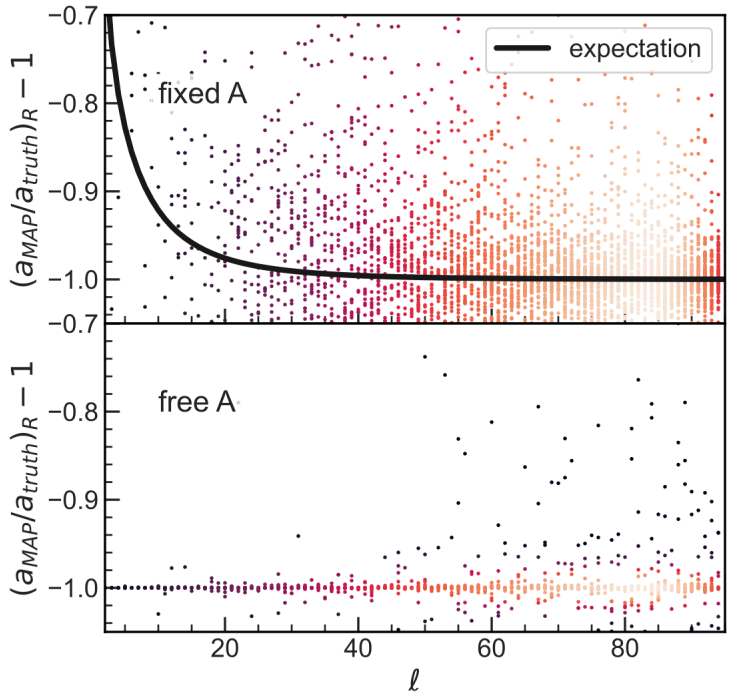

Figure 2.7: The scatter plot of the ratio between the MAP and the truth a_{lm}'s real components. The upper panel shows the case with fixed amplitude and the bottom shows the case with free amplitude. In the case of the fixed amplitude, the mean of the ratios is again the Wiener filter factor (black). In the case of free amplitude, the MAP solution has a null spectrum. Thus the ratios scatter around 0 for all scales.

Now we turn to the reconstructed real space maps (rows 1-3 of Fig. 2.8) which shed more light onto the degeneracy between the reconstructed $s^{\mathcal{P}}$ and $s^{\mathcal{I}}$. Qualitatively, the primordial MAP map captures most features of the true signal, although the small-scale structures are suppressed due to Wiener filtering. However, the MAP estimator completely fails in the ISW reconstruction. In fact, the $\hat{s}^{\mathcal{I},\mathbf{MAP}}$ map looks

like a low-pass filtered $s^{\mathcal{P}}$ map.

The physical explanation is that, when we observe a large-scale hot spot in the sky, it is impossible to confidently associate it with either the primordial CMB or the ISW effect since they both have high amplitudes at low l's. However, when we optimize the posterior function with respect to $s^{\mathcal{I}}$, the algorithm neglects the $s^{\mathcal{P}}$ (Eq. 2.27). Thus, the algorithm inclines to increase the amplitude of the $s^{\mathcal{I}}$ map wherever we observe a large-scale hot spot in $d^{\mathcal{T}}$, even though it is most likely due to $s^{\mathcal{P}}$ since it has a greater power spectrum. As a result, the large-scale hot spots of the reconstructed $s^{\mathcal{I}}$ are heavily correlated with $s^{\mathcal{P}}$, even though they are spatially independent in theory. An analogous bias can be said for the reconstructed primordial map, i.e., the reconstructed primordial map is biased high where there is an ISW hot spot (although it is slightly more difficult to discern in the figure). In short, as we attempt to reconstruct two maps from a single observation using the MAP estimator, the degeneracy introduces significant bias on the field level that correlates with the two reconstructed maps.

Now, we apply the sampler, as defined in §2.3.3, with $s^{\mathcal{P}}, s^{\mathcal{I}}$ as the free parameters. The two-point result is shown in the top panels of Fig. 2.6 in orange, where we observe that the spectra samples scatter around the truth unbiased, and the spectrum of the mean sampled map is equivalent to the MAP spectrum. The sampling result on the field level is shown in rows 1-3 of Fig. 2.8. Here we confirm that the mean sampled map is indeed the field-level MAP solution, and thus suffers from the same bias and degeneracy. Therefore, although the sampling approach solves the multiplicative bias on the two-point level, it is placing the right amount of power in

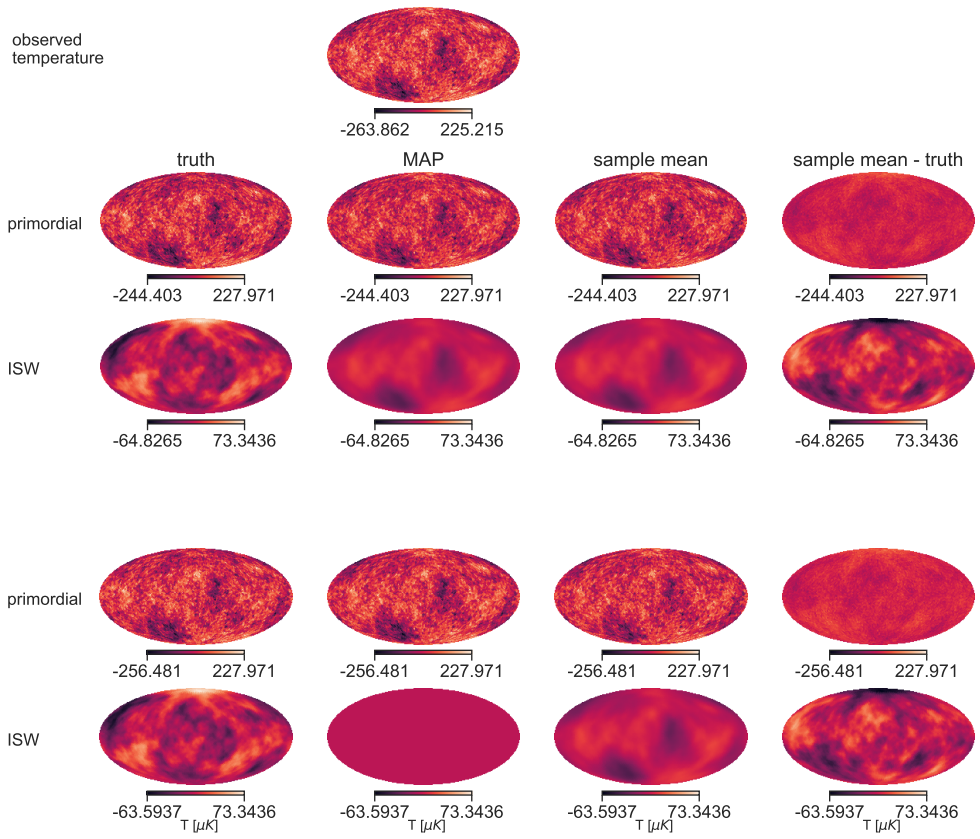

Figure 2.8: Comparison between the observed data (row 1), truth maps, MAP maps, and the mean sampled maps. Rows 2-3 show the model with fixed amplitude, while rows 4-5 show the model of free amplitude. The color bar is shared across each row. Note that in the case of fixed amplitude, the mean sampled map is equal to the MAP map (second and third columns in the second and third rows), while when the amplitude is varied, the two diverge (same columns in the fourth and last row, but most obvious in the ISW maps in the last row) since the posterior marginalized over the amplitude is no longer a simple Gaussian in the signals.

the wrong place at the map level.

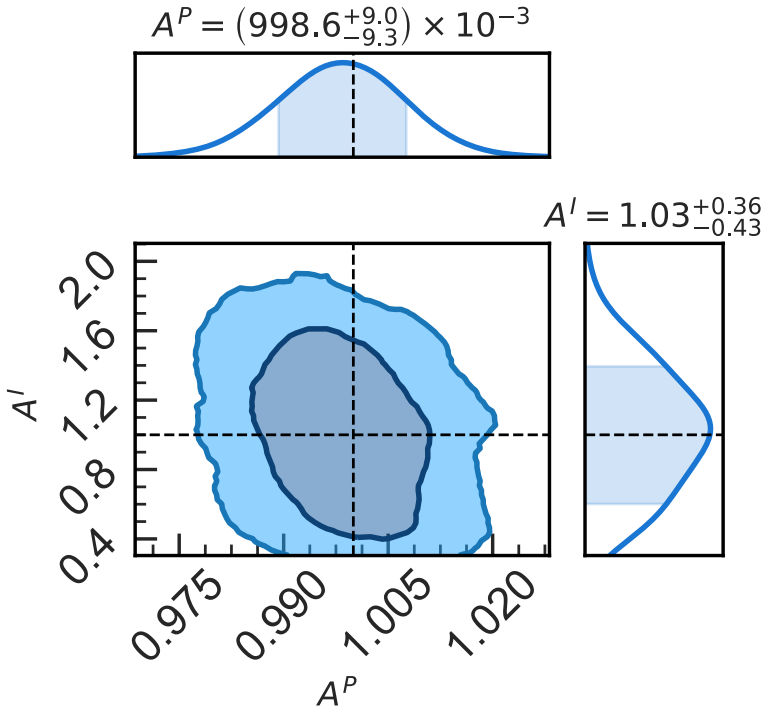

Figure 2.9: The posterior distribution of the primordial and ISW amplitudes with parameter bounds. The data strongly constrain the sum of the two amplitudes (a fact not obvious here because the x-range is so much smaller than the y-range), but the sampling technique manages to correctly infer the values of both.

Varying cosmological parameters

We now vary the two amplitude parameters ($A^{\mathcal{P}}$ and $A^{\mathcal{I}}$) in Eq. 2.25 and attempt to reconstruct them together with the map pixels. The MAP solutions for $\hat{s}^{\mathcal{P},\mathbf{MAP}}$ and $\hat{s}^{\mathcal{I},\mathbf{MAP}}$ are analagous to Eqs. (26) and (27), but with the signal power spectra multiplied by their MAP amplitudes squared, which satisfy

$$\sum_l (2l+1)\left\{1 - \frac{(\hat{A^{\mathcal{P}}}^{\mathbf{MAP}})^2 \mathbb{D}_l \mathbb{C}_l^{\mathcal{P}}}{Q_l^2}\right\} \tag{2.28}$$

$$\sum_l (2l+1)\left\{1 - \frac{(\hat{A^{\mathcal{I}}}^{\mathbf{MAP}})^2 \mathbb{D}_l \mathbb{C}_l^{\mathcal{P}}}{Q_l^2}\right\} \tag{2.29}$$

where

$$Q_l = (\hat{A^{\mathcal{P}}}^{\mathbf{MAP}})^2 \mathbb{C}_l^{\mathcal{P}} + (\hat{A^{\mathcal{I}}}^{\mathbf{MAP}})^2 \mathbb{C}_l^{\mathcal{I}} + \mathbb{C}_l^{n,\mathcal{T}} \tag{2.30}$$

We apply this MAP estimator to the numerical experiment for the CMB-ISW model discussed above. The biased two-point results are shown in the bottom panels of Fig. 2.6 in green. In the case of the ISW reconstruction, the effective noise is so large that the slope of the posterior distribution with respect to the amplitude (Eq. 2.29) never achieves 0. This results in $A^{\mathcal{I}}$ and hence the MAP spectrum being set to 0, which we confirm using direct numerical optimization. This effect is also shown in the bottom panel of Fig. 2.7 where we plot the ratio between the MAP and truth pixel values in harmonic space.

We also construct and apply an HMC-NUT sampler similar to the previous section but with the additional amplitude dependence. The sampled spectra and the spectrum of the mean map are shown in Fig. 2.6 in shaded orange and in red respectively. Similar to the CMB model, the sampled spectra are unbiased. This is also confirmed by the parameter constraint as shown in Fig. 2.9, where we see a 2.9 σ detection of the ISW amplitude. Meanwhile, as the amplitudes are now free, the spectrum of the mean map is still biased, but to lesser degrees than the MAP

solution.

This is consistent with the field-level results (rows 4-5 of Fig. 2.8). Here, for the ISW tracer, the MAP map is essentially constant spatially, whereas the mean map still contains the right amount of power but has placed it all in the wrong place (as in the case of fixed amplitude).

2.6 Joint reconstruction of the CMB, ISW, and the galaxy density maps

We now present the main analysis, where we generalize the framework to jointly analyze data from CMB and wide-field galaxy surveys on the field level. The main goal is the following. Given an observed temperature map and six tomographic galaxy density maps, we want to construct estimates of the primordial ($s^{\mathcal{P}}$), the ISW ($s^{\mathcal{I}}$), and the galaxy density ($\{s^{\mathcal{G},i}\}$) maps, along with two-point and cosmological parameter constraints, all in a consistent and computationally efficient Bayesian framework.

From now on, we will only consider the sampling approach. The addition of galaxy maps introduces off-diagonal terms in the covariance of the posterior distribution, as in Eq. (2.9). As we shall see, following the algorithmic prescription in §2.3.3, we can break the degeneracy between the primordial and the ISW maps using additional maps of the galaxy density.

2.6.1 Theoretical covariance

Equation 2.23 gives the general expression for the theoretical covariance of 2-dimensional tracer fields. In the case of CMB-ISW model, $\langle s^{\mathcal{P}} s^{\mathcal{I}} \rangle = 0$. Now, we introduce tomographic galaxy tracers, which correlate with the ISW effect (but not the primordial CMB) through their common dependence on the matter density field. This correlation will show up as off-diagonal terms in their covariance matrix, as indicated explicitly in Eq. 2.9.

Let $N_i(\chi)$ be the normalized line of sight galaxy density distribution for the redshift bin i, then the galaxy clustering window function that goes into Eq. 2.21 is given by

$$W^{g,i}(\chi) = b_i N_i(\chi) \tag{2.31}$$

where b_i is the linear galaxy bias that connects the matter power spectrum and the galaxy number density power spectrum. Throughout this study, we will treat each b_i as a scale-independent parameter with a fiducial value of 1.

For wide-field photometric surveys, the galaxy redshift distribution N_i varies widely between experiments and catalogs . We consider the MagLim sample from DES Year3 [70, 71] as an example; projected distributions for LSST can be found in [72]. The MagLim catalog consists of 6 redshift bins spanning a redshift range of 0 to 1.2, calibrated using the self-organizing map methods (SOMPZ) and clustering redshifts. The catalog has been extensively tested on simulations and was used by the DES collaboration for the fiducial DES Year3 cosmology analysis [10, 71, 73]. For each redshift bin, we model $N_i(z)$ using the center (z_c) and width (z_w) of the

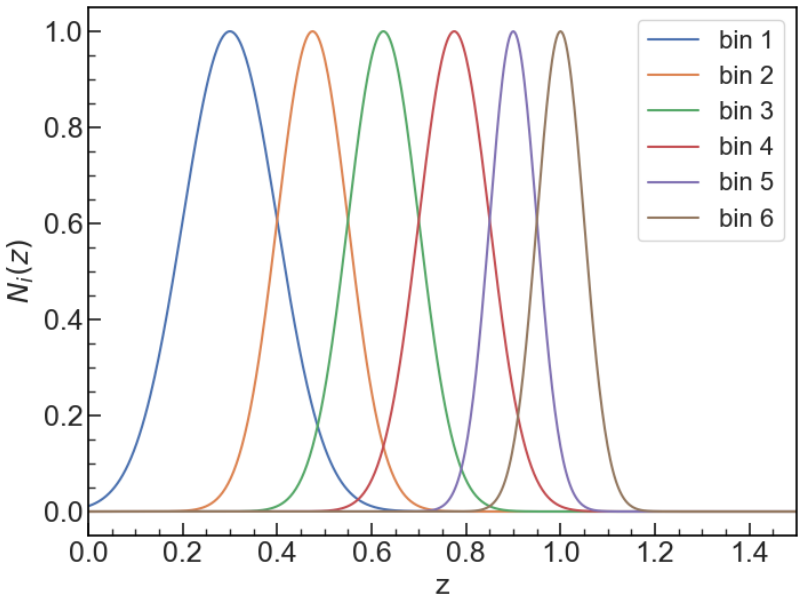

Figure 2.10: The galaxy number densities for each redshift bin as a function of redshift z. $N_i(z)$ is qualitatively modeled on the DES Year 3 MagLim sample.

	bin 1	bin 2	bin 3	bin 4	bin 5	bin 6
z^c	0.30	0.47	0.62	0.78	0.90	1.00
z^w	0.10	0.07	0.07	0.07	0.05	0.05
Σ [deg $^{-2}$]	447.29	319.90	325.48	435.04	316.74	298.85

Table 2.1: The redshift parameters.

distribution, following the functional form

$$\log N_i(z) \propto -\frac{1}{2}\left(\frac{z - z_i^c}{z_i^w}\right)^2 \tag{2.32}$$

We also used the number density (Σ_i) of the MagLim catalog for our simulations but assumed full-sky coverage instead of the DES footprint. We tabulate the binned z_c, z_w, and Σ in Table 2.1 and plot the normalized redshift distribution in Fig. 2.10.

Using the window function in Eq. 2.31, the covariance and the correlation coefficients between the ISW effect and the galaxy density can be computed using Eq. (2.23) and 2.24 (recall that primordial CMB is independent of the other tracers). The correlation coefficients are shown in the lower left corner of Fig. 2.11 in blue. This correlation is the information we hope to leverage to break the degeneracy between the primordial and the ISW maps. The bottom-left panel demonstrates that the ISW effect is most strongly correlated with the galaxy maps on large scales at low redshift.

2.6.2 Cosmological parameters

Let $\mathbb{C}^{\mathcal{P}}$, $\mathbb{C}^{\mathcal{I}}$, and $\mathbb{C}^{\mathcal{G},i}$ be the fiducial power spectra of the primordial, ISW, and the galaxy fields. Similar to the CMB-ISW model, we introduce an amplitude parameter

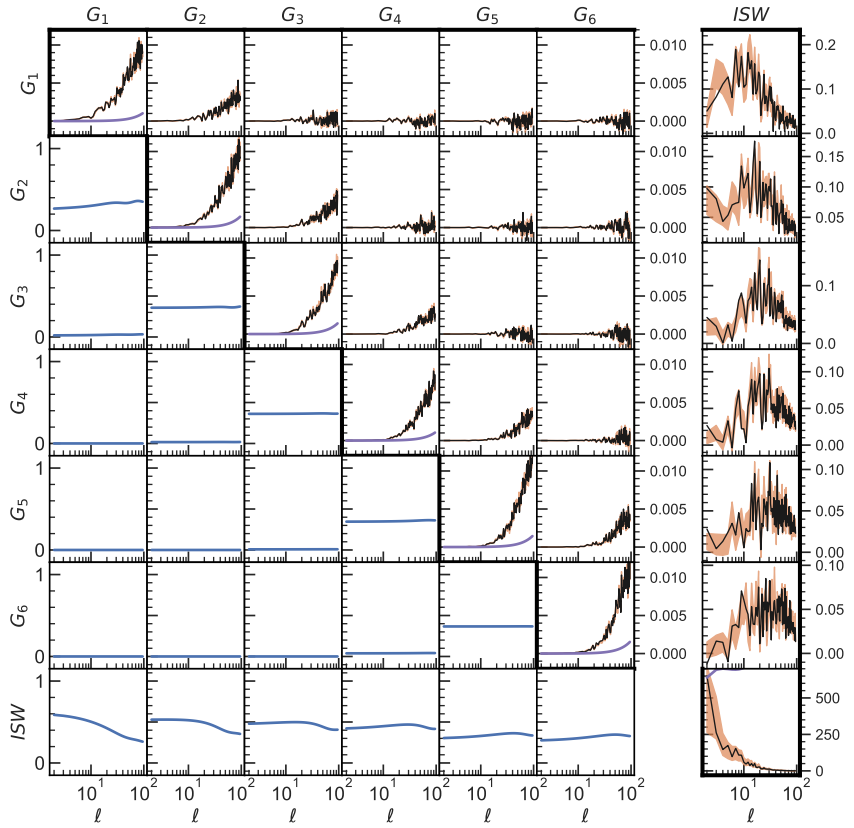

Figure 2.11: The correlation structure between different tracers and the sampler results. The lower-left portion shows the correlation coefficients (blue) between the ISW and the 6 DES Year3 MagLim -like galaxy tracers from different redshift bins. The primordial field is not shown here since it is independent of all late-time tracers. One crucial observation is that (lowest-left panel) the ISW effect correlates most strongly with the low redshift galaxy density maps, since only at late times did dark energy significantly drive the accelerated cosmic expansion. The upper right corner shows the distribution of power spectra sampled from the joint posterior Eq. 2.37 compared to the true power spectra. The input galaxy and ISW signals are shown in black while the distribution of the posterior samples is shown in orange. The diagonal subplots also contain the noise (auto)spectra. Note that the galaxy autospectrum is unitless since the galaxy redshift distribution is normalized. The ISW autospectrum has the unit of μK^2. Because of the different numeric scales, the galaxy-ISW cross-spectra and the ISW autospectrum are broken off into their own column with independent y-axis limits.

for each tracer

$$\mathbb{C}^{\mathcal{P}} \to (A^{\mathcal{P}})^2 \, \mathbb{C}^{\mathcal{P}}$$
$$\mathbb{C}^{\mathcal{I}} \to (A^{\mathcal{I}})^2 \, \mathbb{C}^{\mathcal{I}} \tag{2.33}$$

Equation 2.33 encodes a powerful stress test of ΛCDM . Consider the fiducial growth function $D(z)$. If the truth cosmology deviates from ΛCDM , then the actual growth function will be $A(z)D(z)$, where $A(z)$ is some redshift-dependent factor. Thus, we can interpret $A^{\mathcal{I}}$ as an integral of $A(z)$ over the ISW window function. Then, any detection of $A^{\mathcal{I}} \neq 1$ implies a deviation from the ΛCDM model. The ISW window function is rather wide.

The tomographic galaxy density window function is much narrower. Therefore, multiplying each binned galaxy spectrum by an amplitude factor $A^{\mathcal{G},i}$ would inform us of the consistency of ΛCDM at each redshift slice. This would be, and ultimately will be, a much more strenuous test of the ΛCDM model. Unfortunately, since we consider here only galaxy maps, $A^{\mathcal{G},i}$'s are entirely degenerate with the b_i's in Eq. 2.31. So in this study, we will set $A^{\mathcal{G},i} = 1$. In future works, one could jointly analyze galaxy and shear maps to break this degeneracy and directly constrain the amplitude of the growth function in relatively narrow redshift intervals.

2.6.3 Noise model and simulation

Since we are mostly interested in extracting the large-scale ISW signal, we again perform the simulation on a HEALPix grid of `NSIDE` = 32. We generate the tem-

perature and galaxy tracers using the full covariance as described in §2.6.1. For the observed temperature map, we inject white noise with a variance of 200 μK^2. For the tomographic galaxy maps, we assume a noise spectrum of

$$\mathbb{C}^{n,\mathcal{G},i}(l) = \frac{\pi^2}{180^2 \Sigma_i} \tag{2.34}$$

where Σ_i is the number density of bin i per square degrees. The realized spectra (black) and the modeled noise spectra (purple) are shown in Fig. 2.12 (for the primordial and the ISW signals) and the upper right panels of Fig. 2.11 (for the ISW and the galaxy signals).

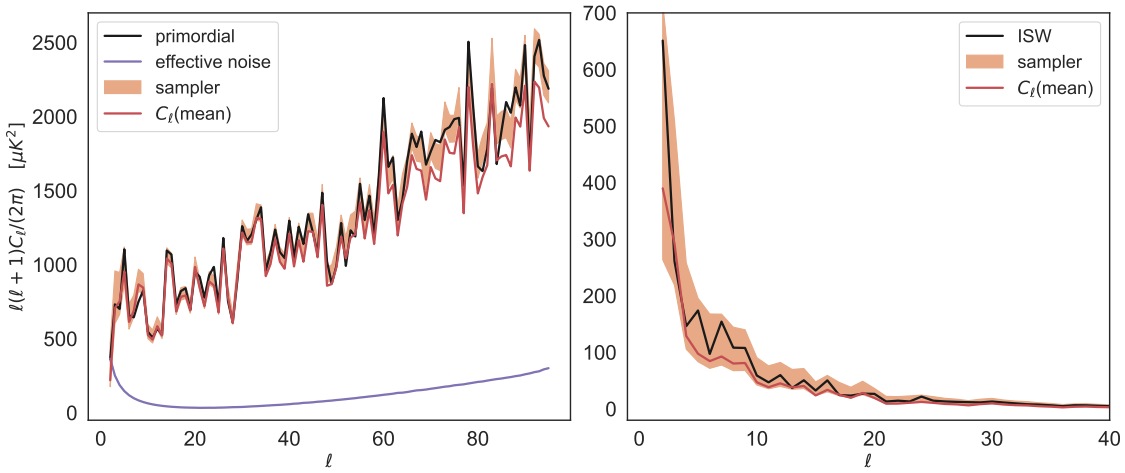

Figure 2.12: The recovered $s^{\mathcal{P}}$ and $s^{\mathcal{I}}$ spectra using the joint reconstruction technique described in §2.6. For each case, the truth spectrum is shown in black, the 1σ credible interval of the sampled spectra is shown in shaded orange, and the spectrum of the mean map is shown in red. The effective noise of the primordial maps is $\mathbb{C}^{\mathcal{I}} + \mathbb{C}^{n,\mathcal{T}}$ shown in purple. The effective noise of the CMB map ($\mathbb{C}^{\mathcal{P}} + \mathbb{C}^{n,\mathcal{T}}$) is above the ISW spectra on all scales and hence not shown in the plot. Notice that both tracers are reconstructed with much higher SNR comparing to the CMB-ISW model.

2.6.4 Posterior modeling and sampling

The observables are

$$d^{\mathcal{T}} = s^{\mathcal{P}} + s^{\mathcal{I}} + n^{\mathcal{T}} \tag{2.35}$$

$$d^{\mathcal{G},i} = s^{\mathcal{G},i} + n^{\mathcal{G},i} \tag{2.36}$$

Therefore, the inference problem is specified by the posterior distribution

$$\begin{aligned} & p(s^{\mathcal{P}},\{s_i^{\mathcal{G}}\}, s^{\mathcal{I}}, A^{\mathcal{P}}, A^{\mathcal{I}}, \{b_i\}|d^{\mathcal{T}}, \{d^{\mathcal{G},i}\}) \qquad (2.37) \\ & \quad \propto \left(\det \left[\mathbb{C}^{n,\mathcal{T}} \prod_{i=1}^{6} \mathbb{C}^{n,\mathcal{G},i} \right] \det \left[\mathbb{C}_l(b_1, ..., b_6, A^{\mathcal{I}}, A^{\mathcal{P}}) \right] \right)^{-\frac{1}{2}} \\ & \quad \times \exp \sum_{lm} \frac{-|d_{lm}^{\mathcal{T}} - s_{lm}^{\mathcal{P}} - s_{lm}^{\mathcal{I}}|^2}{2\mathbb{C}_l^{n,\mathcal{T}}} \\ & \quad \exp \sum_{i=1}^{6} \sum_{lm} \frac{-|d_{lm}^{\mathcal{G},i} - s_{lm,i}^{\mathcal{G}}|^2}{2\mathbb{C}_l^{n,\mathcal{G},i}} \\ & \quad \exp \sum_{lm} \frac{-|\{s_{lm,1}^{\mathcal{G}}, ..., s_{lm,6}^{\mathcal{G}}, s_{lm}^{\mathcal{I}}, s_{lm}^{\mathcal{P}}\}|^2}{2\mathbb{C}_l(b_1, ..., b_6, A^{\mathcal{I}}, A^{\mathcal{P}})} \end{aligned}$$

where $\mathbb{C}_l$ is given in Eq. (2.9) and the notation $|x|^2/\mathbb{C}$ represents the quadratic form $x^T\mathbb{C}^{-1}x$

The first determinant terms encode the noise covariance, which is fixed in our example. The $\mathbb{C}_l$ determinant encodes the posterior's dependence on the tracer amplitudes. The first two exponential terms come from the likelihood of the temperature and the galaxy maps. The last exponential term is the Gaussian priors on

the signals, including the off-diagonal covariance matrix from Eq. (2.9). Not shown here is that all the free amplitudes have flat priors in the interval $[0.2, 3]$.

The structure of the sampler is much the same as in the CMB and the CMB-ISW models, following the prescription of §2.3.3. Notice that since now the covariance $\mathbb{C}_l(b_1, ..., b_6, A^{\mathcal{I}}, A^{\mathcal{P}})$ is nondiagonal and extremely high dimensional, we have to employ the block diagonal Cholesky decomposition method introduced in §2.3.3 to make the sampler computationally feasible.

Despite the high dimensionality of the problem, we find that the chain equilibrates quickly (more details are given in Appendix 2.8.4). In general, the amplitude parameters have a much longer correlation length than latent map parameters during the sampling phase. Among the amplitude parameters, the ISW amplitude has a much longer correlation length of 10^2 samples. Overall, the sampler is very fast. The entire analysis took less than 10 hours on a single Apple M1 CPU.

2.6.5 Results

For each iteration of the sampler after burn-in, we collect a set of maps and parameters

$$\{\mathbf{s}_i, \mathbf{A}_i\} = \{s^{\mathcal{P}}_{lm,i}, s^{\mathcal{I}}_{lm,i}, \{s^{\mathcal{G}}_{lm,k}\}_i, A^{\mathcal{P}}, A^{\mathcal{I}}, \{b_k\}_i\} \tag{2.38}$$

for $k = 1, ..., 6$ labeled by a common sample index i. The collection $\{\mathbf{s}_i\}$ forms the set of posterior samples which we will now analyze. We will present our findings in three parts: cosmological parameter constraints, power spectra reconstruction, and field-level reconstruction. The final results for this joint analysis are shown in Figs. 2.11 - 2.13.

Parameter	Constraint
$A^{\mathcal{P}}$	$1.0029^{+0.0085}_{-0.0089}$
$A^{\mathcal{I}}$	$1.03^{+0.17}_{-0.15}$
b_1	$1.0000^{+0.0070}_{-0.0084}$
b_2	$0.9882^{+0.0083}_{-0.0080}$
b_3	$1.0068^{+0.0081}_{-0.0082}$
b_4	$0.9900^{+0.0083}_{-0.0074}$
b_5	$1.0052^{+0.0081}_{-0.0079}$
b_6	$1.0060^{+0.0080}_{-0.0094}$

Table 2.2: The cosmological parameter constraints given by the CMB-ISW-galaxy model. Here $A^{\mathcal{P}}$ and $A^{\mathcal{I}}$ are the amplitudes of the primordial and the ISW power spectra and b_i are the tomographic galaxy biases.

Cosmological parameter constraints

The constraints on the two temperature tracer amplitudes and the six galaxy biases are shown in Fig. 2.13 and summarized in Table 2.2. All 8 parameters are unbiased within 2σ. Consistent with our previous findings, the best-constrained parameters are the primordial amplitude and the tomographic galaxy biases which all have small effective noise. For these parameters, we achieve percent-level constraints assuming our very simple problem setups. We find that $A^{\mathcal{I}}$ is also unbiased and constrained to around 15% (or a 6.9σ detection), improving dramatically compared to the $\sim 40\%$ constraint (or 2.9σ detection) in the absence of galaxy data.

Power spectra constraints

Besides obtaining the correct overall power spectra amplitudes, we show that the reconstructed power spectra are unbiased for all tracers for all scales. The results for

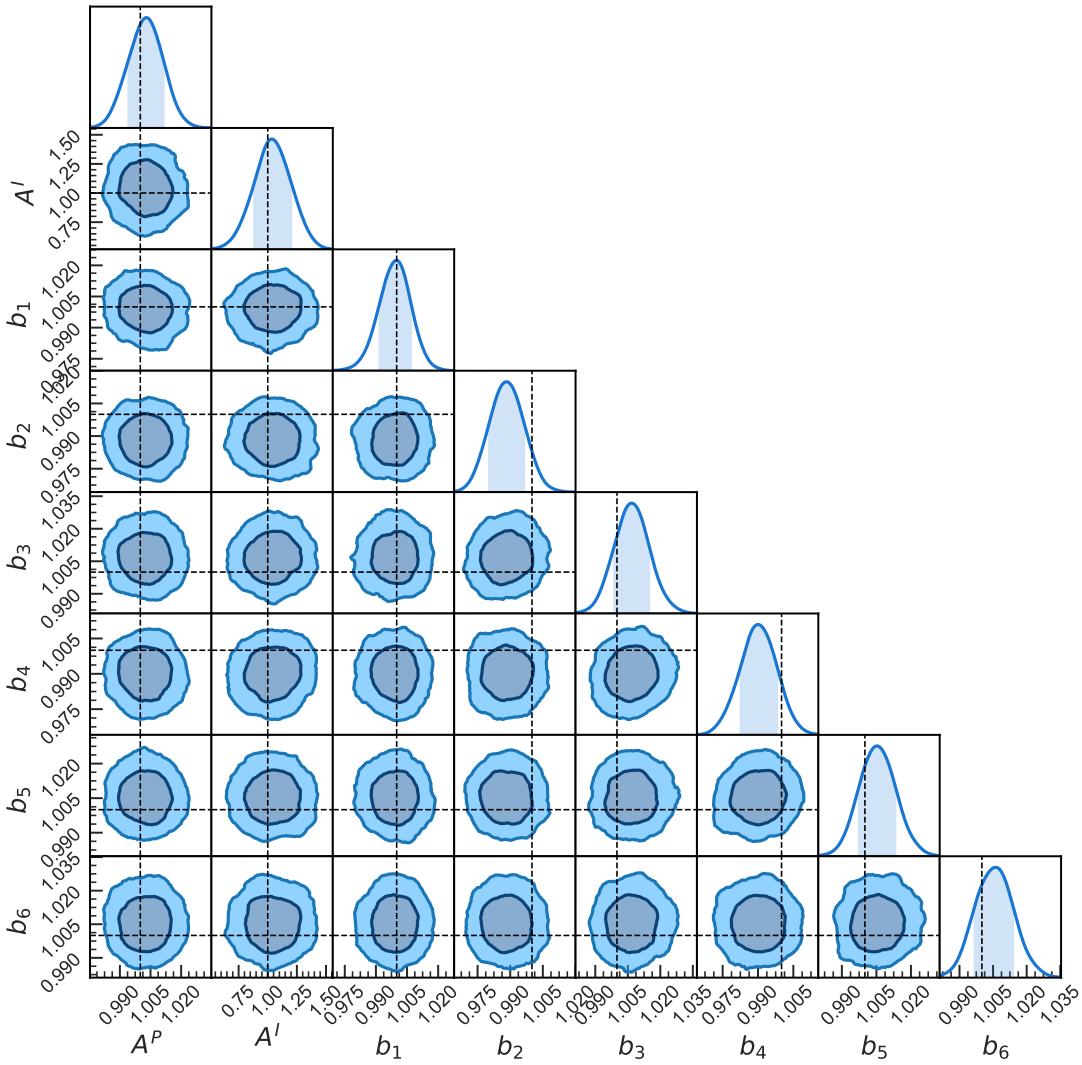

Figure 2.13: The posterior distribution of the amplitudes $A^{\mathcal{P}}$ and $A^{\mathcal{I}}$ and the 6 galaxy biases b_i's. The 2-dimensional contours label the 0.68 and 0.95 credible intervals. The KDE-smoothed histograms above the contours are the parameters' marginal distributions, where the shaded intervals represent the 0.16 and 0.84 quantiles. The truth values are indicated by dashed black lines.

the primordial CMB and the ISW effect are shown in Fig. 2.12, where the sampling result is shown in shaded orange and the truth is shown in black. For ISW, we further observe that the uncertainty of the power spectrum estimation also shrinks considerably around the truth compared to the CMB-ISW model. This gain in SNR is directly attributed to the new information from the galaxy tracer fields.

The sampling result for the ISW and the galaxy tracers (all the tracers that are correlated with each other) are shown in the upper right panels of Fig. 2.11 in orange. Here we see that the method has captured all the auto and cross-spectra of the tracer fields. The galaxy power spectra are especially well-reconstructed, in part due to their intrinsic high SNR observations.

Field-level reconstruction

The CMB-ISW-galaxy model reconstructs tracer maps at higher SNR than previous models. The result for the temperature tracers is shown in Fig. 2.14. Comparing to the CMB-ISW model (Fig. 2.8), we see a dramatic improvement in the reconstruction accuracy. Under the multiprobe joint reconstruction framework, the field-level information in the galaxy maps funnels into the temperature map-making process and efficiently breaks the degeneracy between the primordial and the ISW field. Most notably, although the ISW signal is by far noise dominated on all scales, the mean sampled ISW field is now actually tracing the structures of the true ISW field and decorrelated with the true primordial field. The primordial reconstruction also receives the same benefit, as the residual error of its reconstruction is visibly less correlated with the true ISW field, compared to the CMB-ISW model.

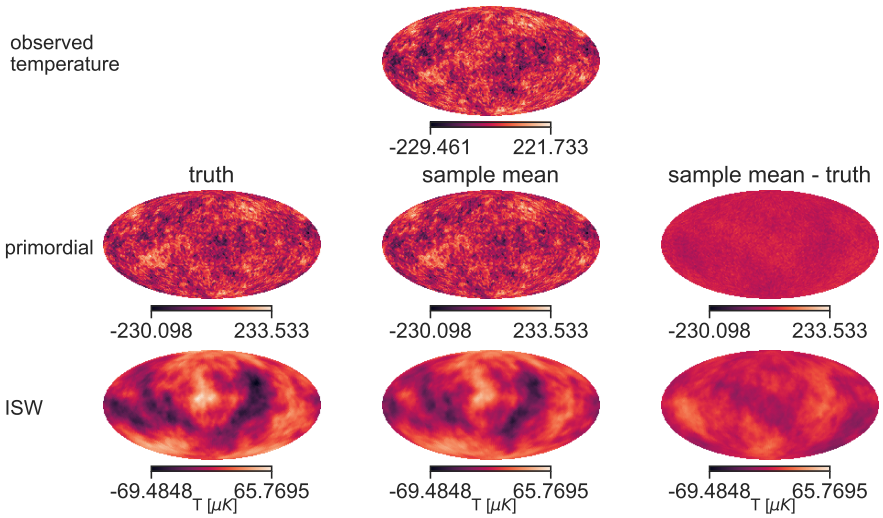

Figure 2.14: Comparison between the observed temperature (row 1), the primordial CMB maps (row 2), and the ISW maps (row 3). For both the primordial CMB and the ISW maps, we compare the truth, the mean sampled, and the difference maps obtained from the CMB-ISW-galaxy model with free cosmological parameters.

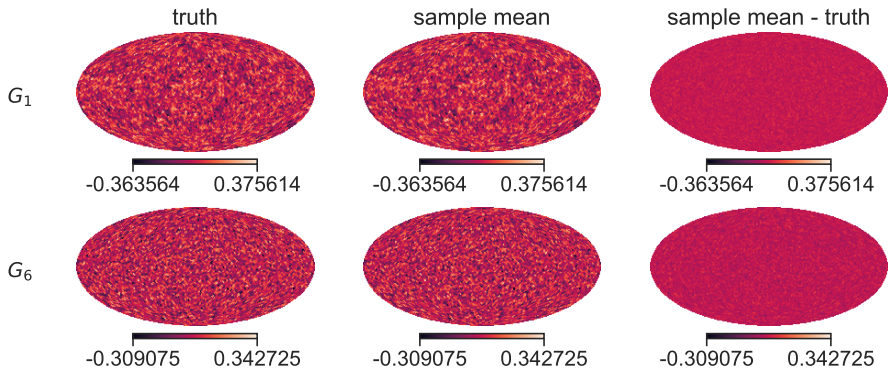

Figure 2.15: Map-level reconstruction of the galaxy fields. Here we only show the first and the last redshift bins, the results for other bins are similar. Unit is in dimensionless overdensity. Note that although the two redshift bins have the same map resolution, the pixel scale of the lower redshift bin corresponds to smaller physical scale. Thus, the lower redshift bin has higher map-level variance, as expected.

The final result also includes samples of denoised tomographic galaxy maps as shown in Fig. 2.15. In theory, the cross-correlation with the ISW component of the temperature map can boost the SNR of the reconstructed galaxy maps as well. However, since the observed galaxy maps are already high in SNR and the degrees of freedom in the galaxy maps by far overwhelm that of the temperature tracers, their improvement is negligible.

2.7 Conclusions

We have implemented a general hierarchical Bayesian framework that employs HMC to sample directly from the joint posterior of the field-level multiprobe model. We did this in a pristine framework: simulated all-sky maps with simple noise properties.

One goal of this is to understand the advantages and limitations of field-level analyses prior to including more realistic effects. The other is more specific: to assess how accurately surveys can measure the ISW amplitude.

We particularly focused on comparing two approaches: maximum *a posteriori* (MAP) and sampling. This enabled us to demonstrate both the well-known bias of the MAP two-point estimator [61] (e.g., the lower jagged curves in Fig. 2.2) and the multiplicative bias of the field-level values (Fig. 2.3). These MAP biases persist as we added more complexity to the data vector and the contributing signals. The sampling approach is also biased at the field level (although, in general to a lesser extent) but is unbiased for power spectra and cosmological parameter constraints, as illustrated in Figs. 2.4, 2.9, and 2.13. This suggests that *the Bayesian posterior sampler produces unbiased cosmological parameters when multiple surveys are analyzed jointly.* Given the potential biases involved in map making and then cross-correlating, this seems to us to provide an excellent justification for the use of field-level analyses

Our results addressing the second goal can be expressed in a single number. Using only CMB data, the amplitude of the ISW signal can be extracted from the CMB only with little power, perhaps at the $2-3\sigma$ level. However, when galaxy survey data is added, we project a 6.9σ detection. Before exploring the limitations of this projection, it is worth emphasizing that the long-term goal is to stress-test ΛCDM, and this method provides a test at the 10-15% level. However, this is but one of a slew of amplitudes that can be measured with upcoming data so we are optimistic about this general idea of constraining amplitudes of large-scale power spectra as a powerful way of testing the fiducial cosmological model.

Back to the limitations of our analysis: assuming full-sky slightly overstates the capability of the CMB, which is masked in the Galactic plane and overstates by at least a factor of two the coverage of, e.g., LSST. Statistically, then, one might reasonably inflate our projections by $\sqrt{2}$. However, there are a number of signals that we did *not* include: galaxy shapes and CMB lensing [22, 51, 52]. The kernels for both of these – especially the former – overlap significantly with that of ISW, so we expect that including them will quite likely recover this factor of $\sqrt{2}$. By adding these observables into the analysis, we could construct tomographic (convolved with different, albeit overlapping, kernels) maps of the matter density in a consistent Bayesian framework.

However, before turning to real data, we must relax the simplifications of our simulations, so we spend the rest of this conclusion alerting ourselves and our readers to those hurdles.

2.7.1 General cosmological parameters

There is no conceptual barrier to including cosmological parameters that modify the shape of the cross-spectra (in contrast to A and b which modulate only their amplitudes). The main challenge is that we must specify the derivatives of the posterior distributions with respect to each of the cosmological parameters in a computationally efficient fashion (e.g., finite difference methods will be too inefficient in an inference algorithm of this scale). The process of differentiating through the Boltzmann code and the Limber approximation computation is especially difficult. We see a few ways for future projects to tackle this issue.

1. Use automatic differentiation to take gradients through the cosmological dependence [74].

2. Use simple fitting functions (e.g., [75]) where the analytical derivatives are easily attainable.

3. Train a neural network-based cosmological emulator where the network is by definition differentiable [76, 77].

4. Exclude cosmological parameters from the HMC sampling altogether. Instead, we can sample the cosmological parameters and the maps iteratively in the Gibbs sampling paradigm [22, 38, 78].

2.7.2 Masking, anisotropic noise, and other systematic effects

The likelihood model we used in this study is very simple. We considered only the case of signal reconstruction on the full sky with isotropic noise and no masking. In order to adapt this algorithm for real-data analysis, we must take into account the limitation of survey geometry for both experiments, as well as foreground and point source masks. Further, the noise in real data will often be anisotropic and often a parametric function of a set of systematics spatial templates. Both the masking and the anisotropic noise models will introduce off-diagonal terms in the covariance matrix computation in Fourier space. Thus, a computationally efficient solution is to still sample the full sky, unmasked, and noiseless maps in Fourier space, transform the maps into real space, and define the likelihood there. Once in real space, we can implement different anisotropic and parametric noise models, and even attempt to

constrain nuisance noise model parameters during sampling as well.

Systematic effects such as foregrounds and survey properties can be handled in the general framework of this field-level analysis. In particular, the *signal* vector can be expanded to include these. This one-step approach – as opposed to the current treatments – may be necessary for future surveys with increased statistical precision. One simple way to understand why this may be needed is that a sample of cosmological parameters that predicts large clustering is more likely to label an ambiguous object a galaxy (rather than a star) if it is near another galaxy. Other systematics – such as photometric redshift uncertainty – can be included by introducing nuisance parameters.

2.7.3 Smaller scales

We have included only large scales here, and there is a huge advantage to doing so, in that the prior distributions of the signals is known to be Gaussian. There is a huge disadvantage to throwing out all the information available on small scales. Including small scales in the posterior requires a knowledge of the prior distribution of the signal, a distribution that is less and less Gaussian as we push to smaller scales. There are two possible approaches to this: (i) assume a simple distribution (e.g., Gaussian or log-normal [76]) and investigate the potential biases by running the pipeline on simulations and (ii) the ambitious approach of rolling the clock back and using the primordial fields as the parameters given the observed highly processed fields (e.g., [79, 80]).

2.8 Appendix

2.8.1 Distributions of individual modes of the sampled power spectra

Here we present the marginal posterior distribution of $\mathbb{C}$ for each individual l mode for the CMB model ($\mathbb{C}^{\mathcal{P}}$, Fig. 2.16), CMB-ISW model ($\mathbb{C}^{\mathcal{I}}$, Fig. 2.17), and the CMB-ISW-galaxy model ($\mathbb{C}^{\mathcal{I}}$, Fig. 2.18).

Figures 2.16 and 2.17 demonstrate that the MAP power spectra amplitudes (vertical blue and orange lines) have greater bias 1) at higher l-modes where the noise power is larger 2) when the power spectra amplitudes are set free. Figure 2.17 and 2.18 show that, in the case of free amplitude parameters, the quality of $\mathbb{C}^{\mathcal{I}}$ reconstruction (the width of its marginal distribution) improves dramatically when one introduces galaxy information.

2.8.2 MAP and sampler estimators

Our algorithm generates samples of the posterior. In the text we made two claims about these samples (for the case of fixed cosmological parameter): (i) the mean (two-point and n-point) power spectrum of the samples is unbiased and (ii) in the single component case, the mean map is equal to the MAP map. To demonstrate this, we let $\{s^{\alpha}_{lm,i}\}$ be the set of maps that we have collected during sampling, where i denotes the sample index and α denotes the type of cosmological probe. Let N be the total number of samples.

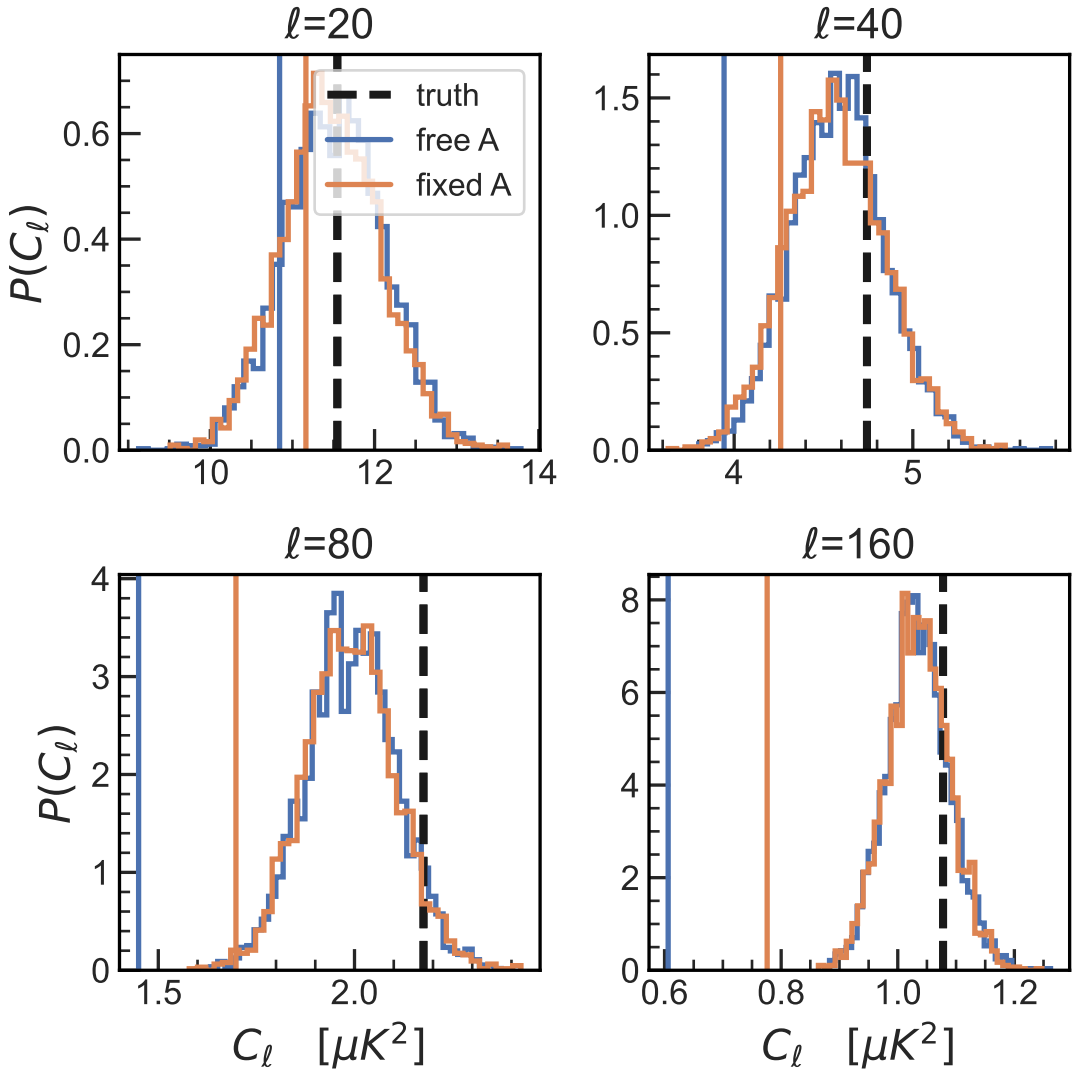

Figure 2.16: The marginal distribution of $\mathbb{C}^{\mathcal{P}}$ of the posterior samples for the CMB model (§2.4). The case with fixed fiducial power spectrum is shown in orange, and the case where one admits a free amplitude is shown in blue. The vertical black lines show the truth $\mathbb{C}^{\mathcal{P}}$, while the colored vertical lines show the MAP values for $\mathbb{C}^{\mathcal{P}}$.

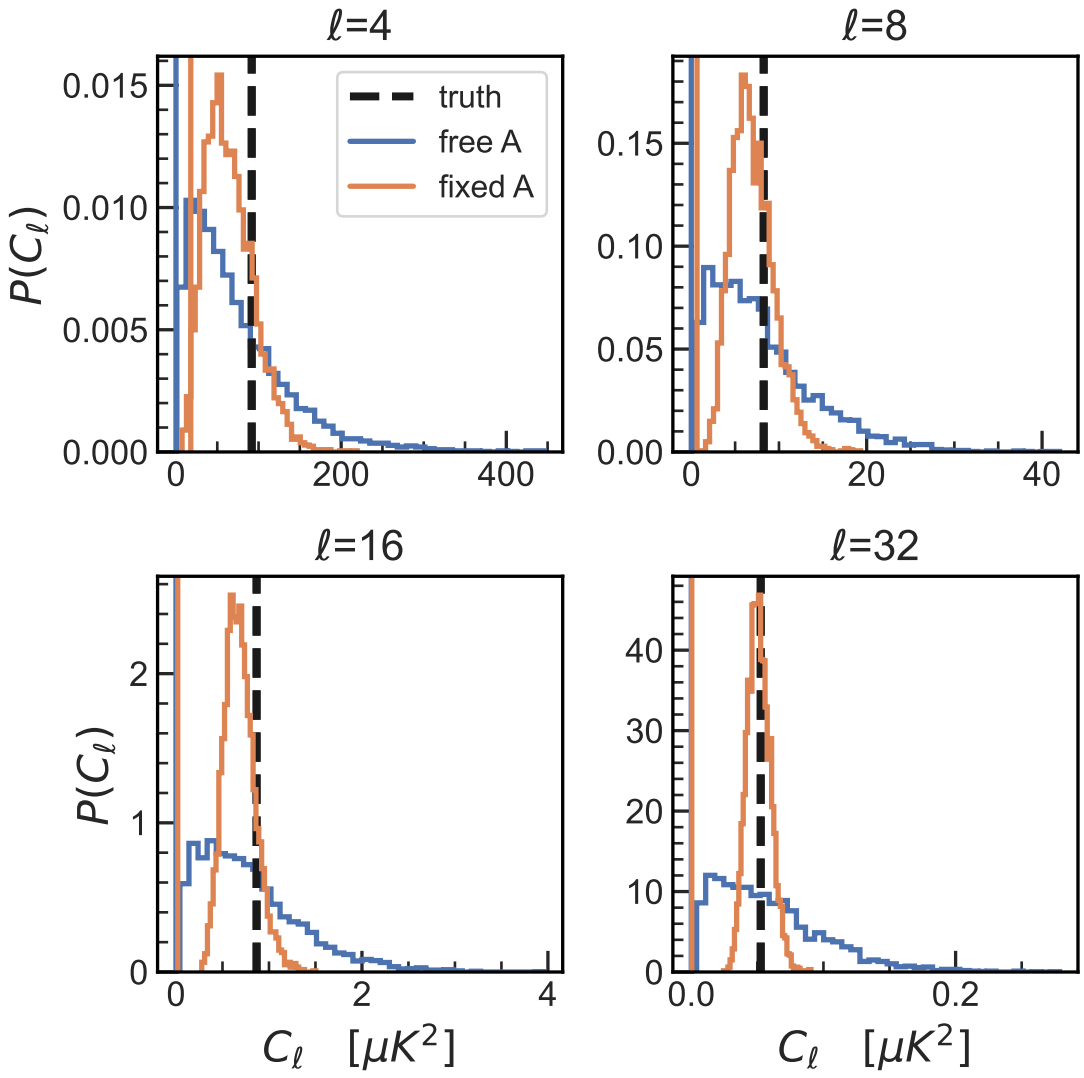

Figure 2.17: The marginal distribution of $\mathbb{C}^{\mathcal{I}}$ of the posterior samples for the CMB-ISW model (§2.5). The case with fixed fiducial power spectrum is shown in orange, and the case where one admits a free amplitude is shown in blue. The vertical black lines show the truth $\mathbb{C}^{\mathcal{I}}$, while the colored vertical lines show the MAP values for $\mathbb{C}^{\mathcal{I}}$.

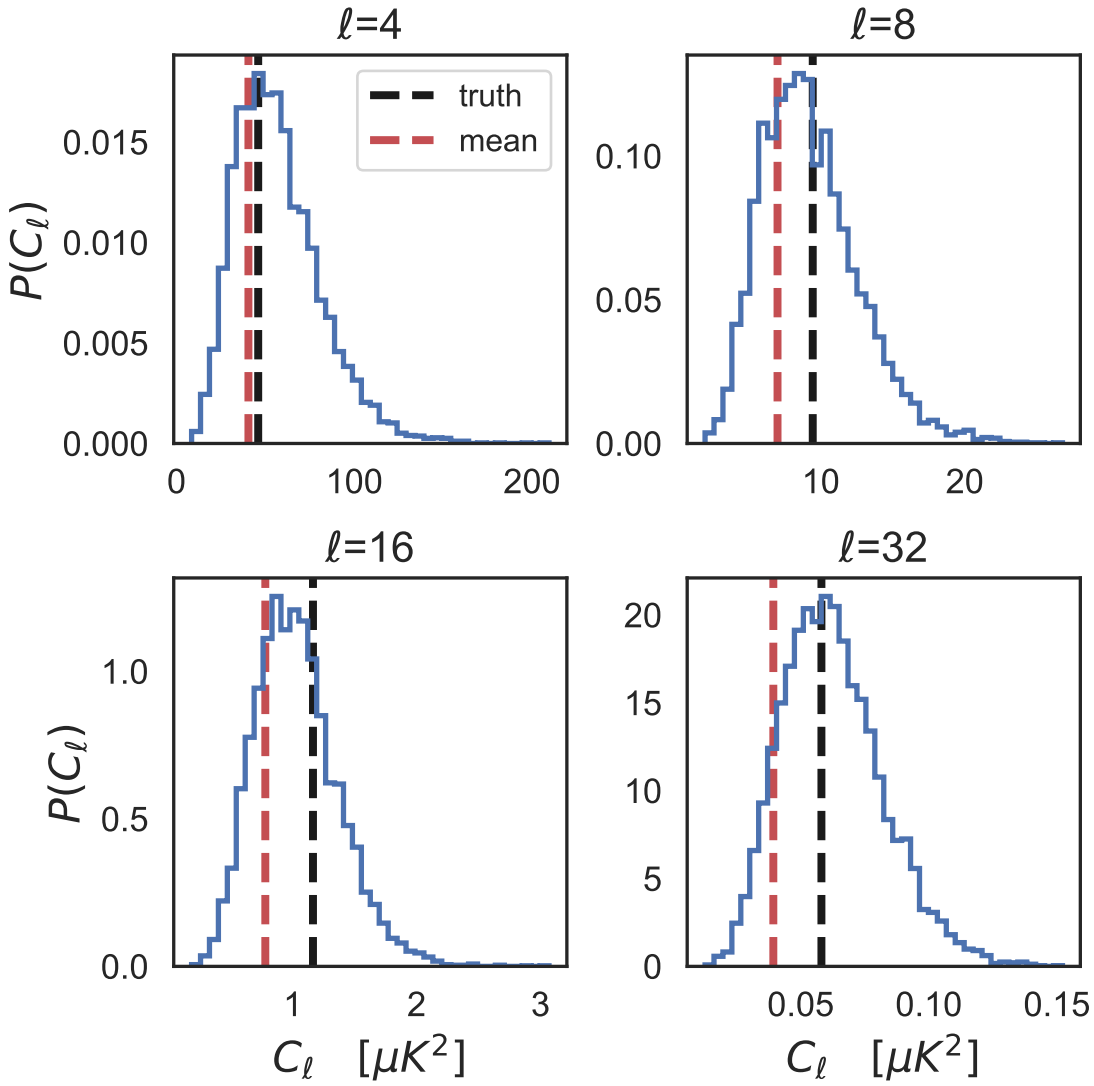

Figure 2.18: The blue lines show the marginal distribution of $\mathbb{C}^{\mathcal{I}}$ of the posterior samples for the CMB-ISW-galaxy model (§2.6). The black and the red dashed lines mark the truth value and the mean of the sample distribution respectively.

Power spectrum

We first show that the mean sampled n-point power spectrum estimator is unbiased. Let us consider the pixel variance, which is the simplest two-point statistic. For a fixed data set D, the mean pixel variance across all the posterior samples is mean($\langle s^2 \rangle$). Here, s is the signal map vector, $\langle s^2 \rangle$ denotes the average pixel variance of s, and the "mean" operation is taken over all the posterior samples. Using i as the index for the posterior samples,

$$\text{mean}(\langle s^2 \rangle) \tag{2.39}$$

$$= \lim_{N \to \infty} \sum_{i=1}^{N} \frac{\langle s_i^2 \rangle}{N} \tag{2.40}$$

$$= \int ds \; |s^2| \; p(s|D) \tag{2.41}$$

where the short hand $|s^2|$ means $\sum_{\alpha=1,\ldots,N_{pix}} s_\alpha^2 / N_{pix}$ with α being the pixel index. Equation 2.41 in general depends on the data D. Practically, this means that mean($\langle s^2 \rangle$) is biased only in the sense that the observed data has intrinsic randomness. If we proceed to integrate over the intrinsic variance in the data, we have

$$\int ds \int dD \; |s^2| \; p(s|D) \; p(D) \tag{2.42}$$

$$= \int ds \; |s^2| \; p(s) \tag{2.43}$$

$$= \langle s^2 \rangle \tag{2.44}$$

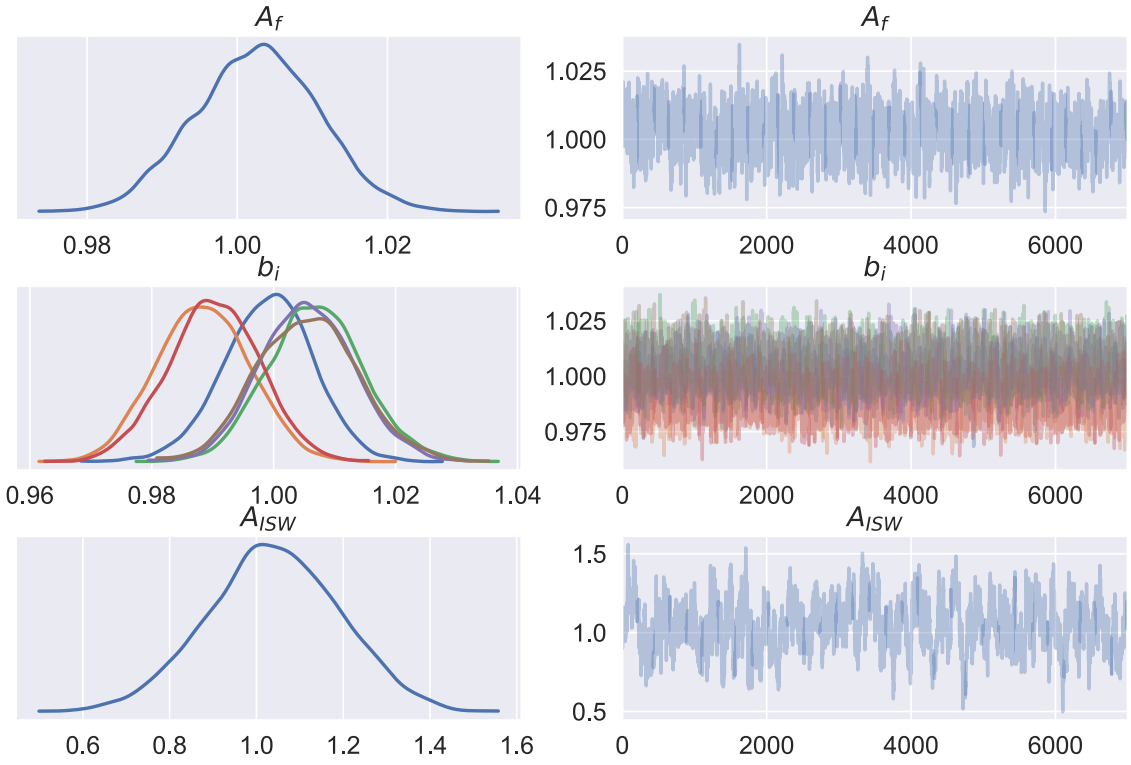

Figure 2.19: The chain convergence of the power spectra amplitudes and galaxy biases for the CMB-ISW-galaxy model. The left panels show the posterior distribution of each parameter after KDE smoothing. The right panels show each parameter's value as a function of sample index after 3000 warmup steps.

which is the true pixel variance the signal. This argument works for the covariance between two different pixels, and more generally, for the n-point correlation functions.

2.8.3 MAP estimator

We claim that the mean sampled map is the field-level MAP solution, and the power spectrum of the mean map is the MAP spectrum. Similar to the argument above,

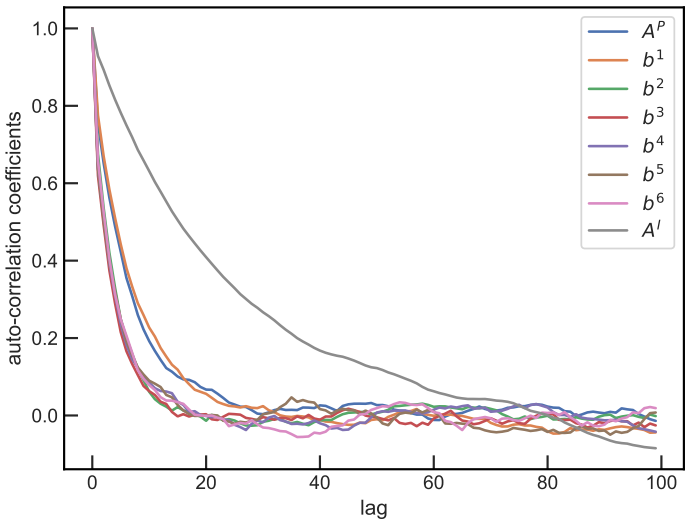

Figure 2.20: The auto-correlation functions of the 8 sampled cosmological parameters shown in Fig. 2.19.

we start with

$$\langle s \rangle = \int ds \; s \; p(s|D) \tag{2.45}$$

where $p(s|D)$ is Gaussian. Since the mean of a Gaussian distribution is also the point of maximum probability, $\langle s \rangle$ is thus also the MAP estimator. More explicitly, by independence between different l, m-modes, we can suppress the l, m subscript on

s and look at the term

$$\text{mean}(s_i) \times p(D) \tag{2.46}$$

$$= \lim_{N\to\infty} \sum_{i=1}^{N} \frac{s_i}{N} \times p(D) \tag{2.47}$$

$$= \int ds \frac{s}{2\pi\sqrt{\mathbb{C}^n\mathbb{C}^s}} \exp\left(-\frac{(D-s)^2}{2\mathbb{C}^n} - \frac{s^2}{2\mathbb{C}^s}\right) \tag{2.48}$$

$$= \frac{\mathbb{C}^s D}{\mathbb{C}^n + \mathbb{C}^s} \times \frac{\exp\left(-\frac{D^2}{2(\mathbb{C}^n+\mathbb{C}^s)}\right)}{(2\pi(\mathbb{C}^n + \mathbb{C}^s))^{1/2}} \tag{2.49}$$

which is the Wiener solution for s times the probability of the data. Dividing by $p(D)$ yields

$$\langle s \rangle = \frac{\mathbb{C}^s D}{\mathbb{C}^n + \mathbb{C}^s} \tag{2.50}$$

as desired. It follows that the power spectrum of the mean map is the Wiener filter spectrum.

Interestingly, notice that the above arguments rests on the symmetry of the Gaussian distribution. If we introduce an additional amplitude parameter A, the distribution over A is no longer Gaussian, and so the mean map is no longer the MAP map solution (as discussed at length in the main text).

2.8.4 Chain convergence

Here we present the chain convergence information of the CMB-ISW-galaxy model, focusing on the 8 cosmological parameters. The sampled parameter values as a function of sample index is shown in Fig. 2.19, and the corresponding autocorrelation

functions are shown in Fig. 2.20. In general, we observe that the map parameters (not shown here) have much shorter correlation lengths than the cosmological parameters. Among the cosmological parameters, $A^{\mathcal{P}}$ and b_i have much shorter correlation lengths compared to $A^{\mathcal{I}}$.

CHAPTER 3

ACCURATE FIELD-LEVEL WEAK LENSING INFERENCE FOR PRECISION COSMOLOGY

3.1 Abstract

We present `Miko`, a catalog-to-cosmology pipeline for general flat-sky field-level inference, which provides access to cosmological information beyond the two-point statistics. In the context of weak lensing, we identify several new field-level analysis systematics (such as aliasing, Fourier mode-coupling, and density-induced shape noise), quantify their impact on cosmological constraints, and correct the biases to a percent level. Next, we find that model misspecification can lead to both absolute bias and incorrect uncertainty quantification for the inferred cosmological parameters in realistic simulations. The Gaussian map prior infers unbiased cosmological parameters, regardless of the true data distribution, but it yields overconfident uncertainties. The log-normal map prior quantifies the uncertainties accurately, although it requires careful calibration of the shift parameters for unbiased cosmological parameters. We demonstrate systematics control down to the 2% level for both models, making them suitable for ongoing weak lensing surveys.

3.2 Introduction

A cosmological field-level inference (FLI) compares the field-level observables (such as galaxy density [20, 24, 46, 81–86], cosmic shear [45, 47, 76, 87, 88], or cosmic

microwave background (CMB) maps [22, 24, 52]) directly to theoretical predictions or simulations. It treats the map pixels and cosmological parameters as random variables to be jointly modeled, with map-making and cosmological parameter inference being carried out simultaneously in a Bayesian manner.

Compared to traditional two-point analyses, FLI has several advantages:

1. In principle, it utilizes all the N-point statistics in the data, so there is minimal information loss

2. It is built around a conceptually simple forward model which easily extends to new physics and systematics

3. It combines the map-making and parameter inference into a single step, enabling joint inference of multiple surveys on the map level

4. It avoids the needs the compute complex high-dimensional covariance matrices

Previous works [88–90] have shown that FLI can achieve superior cosmological constraints compared to two-point analyses. The goal of this work is to establish FLI methods with rigorously demonstrated systematics control at least at the level needed for current imaging surveys.

$$\mathcal{P}(\mathbf{c}, \mathbf{s}|\mathbf{d}) \propto G(\mathbf{d} - \mathbf{s}; 0, N), \times \mathcal{P}(\mathbf{s}|\mathbf{c}), \times, \text{prior}(\mathbf{c}), . \tag{3.1}$$

Here, the set of cosmological and astrophysical parameters are denoted $\mathbf{c}$; the map pixel values (the *signal*, $\mathbf{s}$; the data pixel values $\mathbf{d}$; and $G(\mathbf{x}; \mu, N)$ is a Gaussian distribution of $\mathbf{x}$ with mean μ and variance N. We will treat the noise as uncorrelated

so that the matrix N is diagonal (in the case of lensing, related to the shape noise). The second probability on the right encodes two broad areas of complexity. First, even in the simplest case, where the observations are on the largest of scales so the fundamental fields are Gaussian, care needs to be taken to convert the true field to $\mathbf{s}$, the field that will be compared to the data [91]. We call this set of issues *analysis systematics*, and present our treatment of them in § 3.6 below. Second, most observations are not sampling purely Gaussian fields, so understanding the distribution from which the true fields are drawn is paramount. We call this uncertainty *model misspecification* and discuss this in § 3.7 below. This paper aims to make two main points:

1. We have developed a pipeline to analyze weak lensing data in the flat sky case[1] that includes and quantifies analysis systematics such as pixelization effects, boundary conditions, and general issues that arise when moving between real and Fourier space.

2. In weak lensing, the choice of a Gaussian prior for $\mathcal{P}(\mathbf{s}|\mathbf{c})$ leads to unbiased means on the extracted parameters but incorrect error bars. Meanwhile, using a log-normal prior for $\mathcal{P}(\mathbf{s}|\mathbf{c})$ leads to correct errors bar but the means of the extracted parameters are very sensitive to the exact parametrization of the log-normal distribution.

1. The flat sky case has the advantage of requiring less computational resources and also applicable to deep surveys that cover only a fraction of the sky. Indeed, we will test our methods on realistic mocks that simulate the Hyper Suprime-Cam SSP Survey [92] year-3 shape catalog [93], and the pipeline developed here may be applicable to the Roman Space Telescope [94], current baseline wide-field imaging survey will cover about 4% of the sky.

Looking forward (§ 3.5), we will constrain the tomographic power spectrum amplitudes A_i, which is similar to a tomographically decomposed version of S_8 in standard cosmological analyses. The HSC Year 3 real [95] and Fourier [96] space shear analyses both yield an approximately 4.1% constraints on S_8 ($0.769^{+0.031}_{-0.034}$ and $0.776^{+0.032}_{-0.033}$ respectively). This translates to approximately an 8% constraint on A_i. Since we typically require the systematic uncertainty to be under 1/4 of the total error budget, we set a target of 2% for the absolute bias induced by systematic effects. Given the above numbered points, we chart out a pathway for analyses that overcomes these difficulties at this target 2% level.

This paper is organized as follows. We briefly review the basics of weak lensing in § 3.3, describe the simulated datasets in § 3.4 and then give an overview of the inference pipeline in § 3.5. The first set of results is presented in § 3.6, which deals with analysis systematics, and the second in § 3.7, where model misspecification is discussed. We summarize and outlook in § 3.8.

3.3 Weak lensing basics

3.3.1 Weak lensing maps

As the light from distant galaxies travels toward us, it is deflected by the intervening matter overdensity perturbation δ_m. Consider a line-of-sight (LOS) in the $\boldsymbol{\theta}$ direction and a light source at comoving distance χ. The light source, although emitted at the true position $\boldsymbol{\alpha}$, is observed at $\boldsymbol{\beta}$ due to gravitational lensing. In the weak lensing

regime, the effect of this distortion is approximately linear

$$\frac{\partial \boldsymbol{\beta}}{\partial \boldsymbol{\alpha}}\Big|_{\boldsymbol{\theta},\chi} = \begin{pmatrix} 1-\kappa(\boldsymbol{\theta},\chi)-\gamma_1(\boldsymbol{\theta},\chi) & -\gamma_2(\boldsymbol{\theta},\chi) \\ -\gamma_2(\boldsymbol{\theta},\chi) & 1-\kappa(\boldsymbol{\theta},\chi)+\gamma_1(\boldsymbol{\theta},\chi) \end{pmatrix},, \tag{3.2}$$

where $\kappa(\boldsymbol{\theta},\chi)$ and $\gamma(\boldsymbol{\theta},\chi) = \gamma_1(\boldsymbol{\theta},\chi) + i\gamma_2(\boldsymbol{\theta},\chi)$ are called the convergence and the shear maps [97].

In a tomographic weak-lensing survey, we categorize the source galaxies into redshift bins with normalized galaxy density distributions $n_{\mathrm{gal},i}(z)$. We are interested in the line-of-sight (LOS)-averaged convergence map, which, for a spatially flat universe, is related to the matter overdensity via

$$\kappa_i(\boldsymbol{\theta}) = \int_0^{\chi^*} d\chi W_i(\chi)\delta_m(\chi\boldsymbol{\theta},\chi),, \tag{3.3}$$

$$W_i(\chi) = \frac{3}{2}H_0^2\Omega_m(1+z(\chi))\int_\chi^{\chi^*} d\chi' n_{\mathrm{gal},i}(\chi')\frac{\chi(\chi'-\chi)}{\chi'},, \tag{3.4}$$

where H_0 is the Hubble constant, $\Omega_m = 8\pi G\rho_m/(3H_0^2)$, ρ_m is the matter density, and χ^* is the comoving horizon.

3.3.2 Weak lensing statistics

The n-point correlation functions of the convergence field carry significant cosmological information. The two-point cross-power spectrum between redshift bins i and j is given by (under Limber's approximation [98])

$$C_{ij}(\ell) = \int_0^{\chi^*} d\chi \frac{W_i(\chi)W_j(\chi)}{\chi} P_\delta(k=\frac{\ell}{\chi};\chi),, \tag{3.5}$$

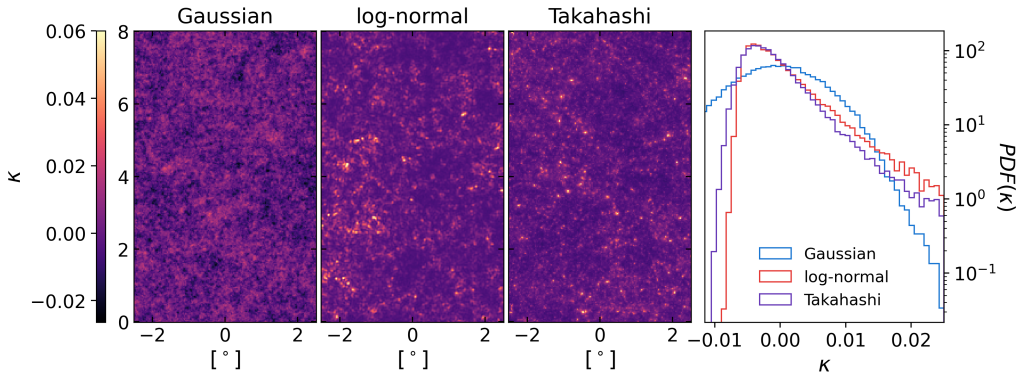

Figure 3.1: Examples of Gaussian, log-normal, and Takahashi fields. All three fields have the same two-point statistics (modeled after the first redshift bin § 3.4) but a visibly different field-level appearance. Their one-point PDFs are shown in the right plot, where we see that the log-normal field has a stronger cutoff at small κ compared to the others. We limit the maximum color map scale to $\kappa = 0.06$ to emphasize the visual distinction between the fields.

where $P_\delta(k;\chi)$ is the 3-dimensional matter power spectrum. In practice, we compute Eq. 3.5 using `camb` [99].

If the convergence field were completely Gaussian, then Eq. 3.5 would contain all the information. In reality, however, small-scale structure formation leads to non-Gaussianity characterized by extended voids and density peaks that contains important cosmological information [100, 101]. For example, Fig. 3.1 shows three examples of convergence fields generated with models (Gaussian, log-normal, and realistic N-body ray-tracing simulation). Although these fields all share the same power spectrum, they are visibly different, each with its own one-point PDF and higher-order statistics. In this work, we model the field-level statistics to extract information not captured by the standard two-point analyses.

One of the most studied models of the convergence PDF is the log-normal distri-

bution [76, 102–104]. Let $\kappa^{\rm G}$ be a zero-mean Gaussian (transfer) field. The associated zero-mean log-normal field is defined by applying a transformation across the entire Gaussian field,

$$\kappa = \mathbb{L}(\kappa^{\rm G}, a) = a\left(e^{\kappa^{\rm G} - {\sigma^G}^2/2} - 1\right),, \tag{3.6}$$

where a is the shift parameter and σ^G is the standard deviation of the transfer map. The log-normal field is fully characterized by its variance σ^2 and shift parameter, giving it one more degree of freedom than the Gaussian field. The shift parameter a is particularly interesting, as it corresponds to the minimum value of the log-normal distribution. Alternatively, a can be reparametrized in terms of the skewness s [104] by

$$a = \frac{\sigma}{s}\left(1 + y(s) + y^{-1}(s)\right),, \tag{3.7}$$

$$y(s) = \left(\frac{2 + s^2 + s\sqrt{4+s^2}}{2}\right)^{1/3},, \tag{3.8}$$

$$s = \langle(\kappa/\sigma)^3\rangle,, \tag{3.9}$$

which we find to be a more robust measurement on real simulations. The assumption about a is very important and we will discuss it in § 3.7.1.

Finally, we will need to relate the two-point correlation function of the transfer field to that of the log-normal field. This is given by

$$\xi^G_{ij}(\theta) = \log\left(\frac{\xi_{ij}(\theta)}{a_i a_j} + 1\right),. \tag{3.10}$$

The two-point correlation function is related to the power spectrum by the classic Hankel transform

$$\xi_{ij}(\theta) = \mathcal{H}^{-1}\left[C_{ij}(\ell)\right](\theta) \tag{3.11}$$

$$= 2\pi \int d\theta\ \theta\ C_{ij}(\ell)\ J_0(\ell\theta),, \tag{3.12}$$

where J_0 is the Bessel function of the first kind.

3.3.3 Weak lensing observables

A weak lensing survey measures the shapes of galaxies to infer the convergence and/or the shear [105]. For a galaxy with observed complex shear γ_o,

$$\gamma_o = \gamma + \gamma_I + \gamma_n,. \tag{3.13}$$

The first term on the right hand side is the lensing shear, γ_I is the coherent distortion of galaxy shape due to intrinsic alignment (IA) [44, 106], and γ_n is the galaxy shape noise caused by random intrinsic galaxy shape and image noise. In this study, we leave $\gamma_I = 0$ and model γ_n as a zero-mean Gaussian random variable with standard deviation $\gamma_{\rm rms}$. Additionally, higher-order differences between reduced shear and shear [107] are not considered in our analysis.

3.4 Simulations and Datasets

We model our analysis after the HSC Year 3 survey specifications [1, 93]. We use three sets of independent simulations to test our analysis pipeline across the entire galaxy catalog to cosmological constraints process. In this section, we discuss the process of converting galaxy catalog to the observed pixelized shear maps $\gamma_i^{\rm obs}$. We will show in § 3.6 that this procedure introduces significant statistical artefacts in $\gamma_i^{\rm obs}$ that must be reflected in our forward model for unbiased cosmological constraints.

The first simulation comes from Takahashi *et al.* [108], where the authors perform multiple-lens plane ray-tracing on high-resolution N-body simulation. This simulation uses the WMAP 9 years result [9] as its fiducial cosmology, which we also adopt through out our work. The Takahashi simulation has been extensively tested and used in HSC to derive covariance of 2-point summary statics [109]. It provides independent full sky realizations of $\kappa(\boldsymbol{\theta}, \chi)$ and $\gamma(\boldsymbol{\theta}, \chi)$ at the 38 comoving radial shells. These maps are provided in `HEALPix`, format [54, 110] with a resolution of `nside` = 8192,. We use the radial shells to construct the 4 tomographic $\kappa_i(\boldsymbol{\theta})$ and $\gamma_i(\boldsymbol{\theta})$ maps using the HSC Year 3 redshift distribution (Fig. 3.2). Next, we locate rectangular patches of side lengths $(L_x, L_y) = (8^\circ, 5^\circ)$ and the patches are separated by at least 8° to avoid spatial correlations. We generate galaxies with tomographic effective number density ρ_i with true κ and γ values according to their positions.

We set ρ_i to be 36.1, 52.0, 39.6, and 20.6 arcmin^{-2} for the four redshift bins, corresponding to 10 times the HSC Year 3 W12H field specification (explained below). Then, we add Gaussian shape noise to individual galaxies according to Eq. 3.13 with $\gamma_{\rm rms} = 0.26$ [111]. We construct the pixelized maps $\kappa_i^{\rm true}$ and $\gamma_i^{\rm true}$ using the noiseless

catalog, and construct $\gamma_i^{\rm obs}$ using the noisy catalog. In both cases, the pixels have size $\Delta = 3'$.

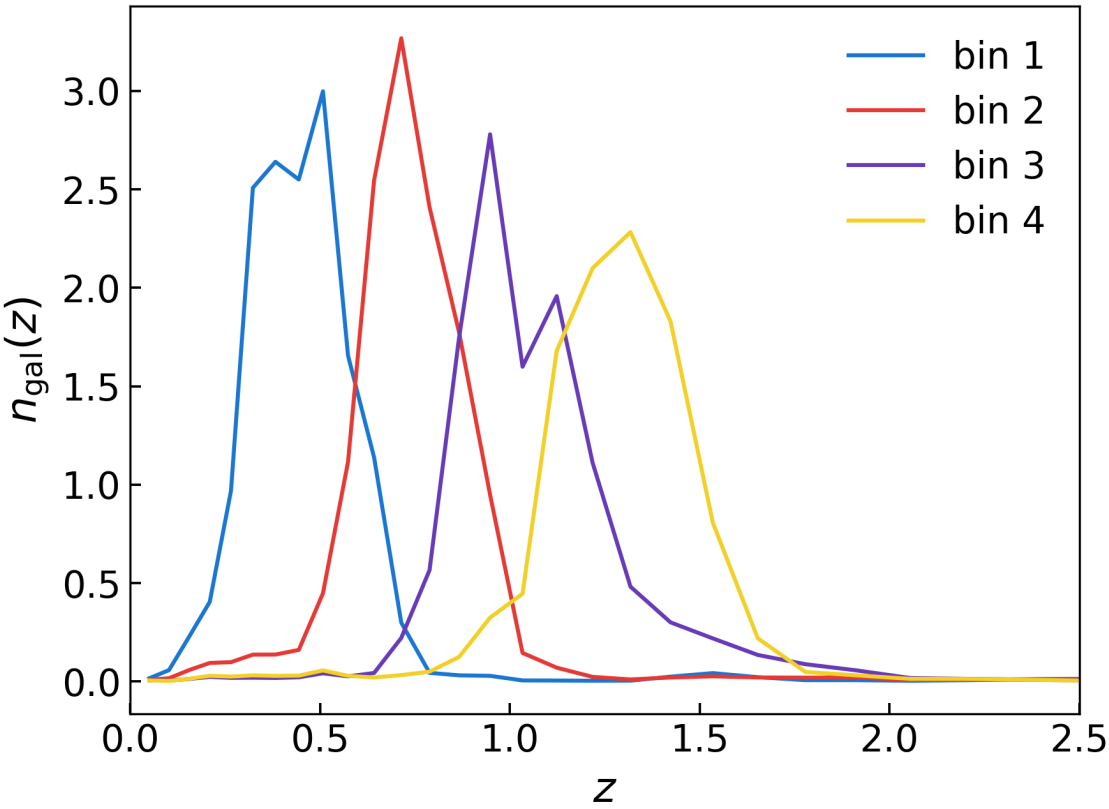

Figure 3.2: The normalized PDF of the tomographic redshift distributions of the HSC Year 3 galaxy shape catalog [1].

The choice of ρ_i merits a few comments. For a pixel p, the effective shape noise has variance

$$N_i(p) = \frac{\gamma_{\rm rms}^2}{\rho_i(p)\ \Delta^2},. \tag{3.14}$$

We are mostly interested in the bias of modeling choices and the robustness of our posterior estimation in the presence of noise and cosmic variance. Thus, given a finite

computational resource, we want to reduce the statistical uncertainty and perform as many independent HSC-like experiments as possible. We achieve this goal by setting ρ_i to 10 times the fiducial HSC number density [93] to reduce the effective shape noise.

Due to simulation artefacts [108], the Takahashi convergence maps does not recover the theoretical power spectrum (Eq. 3.5) with the precision required for this work. Therefore, we need calibrate the theory when performing inference on the Takahashi mocks. The exact procedure is detailed in App. 3.10. The calibrated theoretical power spectrum agrees with the simulation within 1% for $\ell < 24000$ for all redshift bins.

The second and third sets of simulations are conceptually similar. Using the same WMAP9 cosmology, we make full sky log-normal and Gaussian realizations of convergence and shear at resolution of `nside` = 8192, and then construct shear catalogs and flat sky maps as before. We generate the log-normal maps using `FLASK` [104]. For these maps, we observe a larger-than-1% surplus of power above $\ell = 5000$. Therefore, we only use $\ell \leq 5000$ for both the log-normal maps and the log-normal theory power in the analyses.

For all three simulations, we use 4 independent full-sky simulations to generate 80 nonoverlapping flat-sky patches.

3.5 Inference Framework

Our inference framework `Miko`, graphically illustrated in Fig. 3.3, is a hierarchical Bayesian network that forward models from the cosmological parameter to the ob-

served noisy fields. We focus on the lensing convergence and shear fields in this work, but the framework is set up to include galaxy density and CMB fields as well [24]. The weak lensing pipeline includes two models; the first generates the observable fields using a Gaussian map prior, and the second uses a log-normal prior.

`Miko` has four main components. Starting from the top of the flow chart, we first draw cosmological parameters $\mathbf{c}$ and the latent tomographic map parameters $\mathbf{x}_i$ from their prior distributions. For the Gaussian model, we set the free cosmological parameters to be the amplitudes of the tomographic power spectra,

$$C_{ij} \longrightarrow A_i A_j C_{ij}, . \tag{3.15}$$

where C_{ij} is the power spectrum at the fiducial cosmology calibrated to the simulations[2] (§ 3.4). Inference of A_i is a powerful test of the ΛCDM model. Consider the redshift-modulated growth function $A(z)D(z)$. In this context, A_i can be thought of as $A(z)$ integrated over the tomographic window function $W_i(z)$. Thus, a detection of $A_i \neq 1$ implies that the growth history near the redshift interval of bin i deviates from the ΛCDM model. Through out this study, we use a broad and uniform prior on A_i over the interval $[0.5, 1.5]$. For the log-normal model, the free parameters are both A_i and the shift parameters a_i. The prior on a_i is more complicated and we will delay its discussion until § 3.7. The latent map parameters $\mathbf{x}_i$ represent the field-level degrees of freedom and are drawn from the unit normal distribution.

Next, we use the cosmological parameters and survey and analysis specifications

2. In real observations, we can simply set C_{ij} to the prediction of the fiducial cosmology.

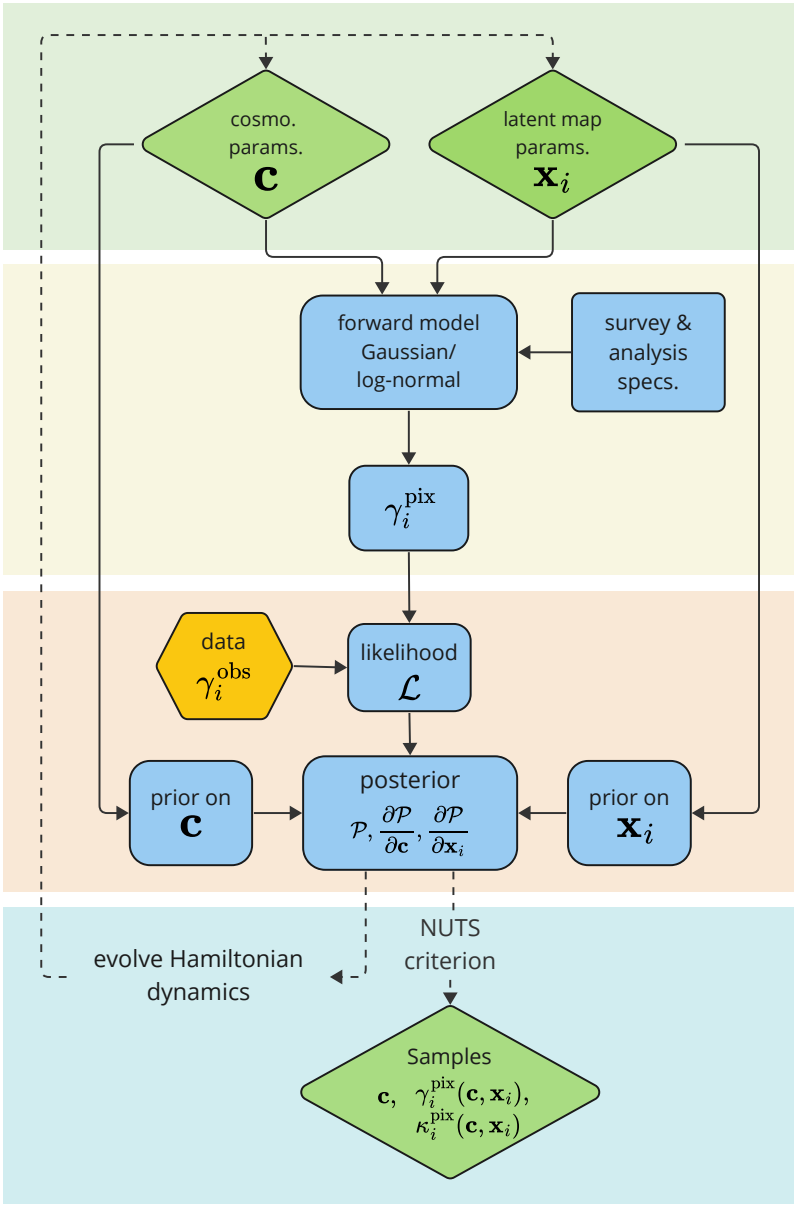

Figure 3.3: This flow chart describes `Miko`'s weak lensing field level forward model and inference pipeline (§ 3.5). The pipeline has four steps (from top to bottom): generating latent parameters, forward modeling the noiseless fields, computing the likelihood given the data, and sampling. In the chart, diamonds represent sampled parameters, squares the model-relevant functions, and the hexagon the observed data vector.

to define a forward model, which transforms $\mathbf{x}_i$ to noiseless and pixelized tomographic convergence and shear maps $\kappa_i^{\rm pix}$ and $\gamma_i^{\rm pix}$ on one (or more) flat sky patch(es). It is important to note that the process of generating $\kappa_i^{\rm pix}$ and $\gamma_i^{\rm pix}$ includes not only the cosmological model but also the survey and analysis-specific systematics model. The process of going from $\mathbf{x}_i$ to $\gamma^{\rm pix}$ is explained in detail in § 3.6). We then compare the noiseless $\gamma_i^{\rm pix}$ to the data $\gamma_i^{\rm obs}$ and build the Bayesian posterior function given by

$$\begin{aligned}
&\log \mathcal{P}\left(\{\gamma_i^{\rm pix}(\mathbf{c}, \mathbf{x}_i)\}|\{\gamma_i^{\rm obs}\}\right) \\
&= -\sum_i \frac{\left(\gamma_i^{\rm obs} - \gamma_i^{\rm pix}(\mathbf{c}, \mathbf{x}_i)\right)^2}{2N} + \log \text{prior}\left(\{\gamma_i^{\rm pix}(\mathbf{c}, \mathbf{x}_i)\}\right), ,
\end{aligned} \tag{3.16}$$

where N is the shape noise variance map given by Eq. 3.14. We implement all computations using differentiable programming in `jax`.

In the last step, we connect this differentiable posterior function to the `numpyro`[59] implementation of the Hamiltonian Monte Carlo No-U-Turn sampler (HMC NUTS) [27, 58, 112] to efficiently sample from the high-dimensional joint posterior space of both the cosmological and the map parameters. The result of the inference is comprised of the posterior samples of the cosmological parameters $\mathbf{c}$ and the maps $\kappa^{\rm pix}$ and $\gamma^{\rm pix}$. We could also obtain samples of the continuous fields (without the pixelization effect) if desired.

We perform consistency checks on both models. This is done by first using the model to generate noisy shear maps and using the same model to infer A_i from it. We use 10 chains for each experiment for all the main analyses in this paper. Each

chain is initialized at the maximum a posterior estimate calculated with a stochastic variational inference procedure parametrized with the δ-distributions. We then run approximately 500 warm-up NUTS steps followed by 100 sample collection NUTS steps. We use a target probability of 0.7 and a maximum tree depth of 9 for our NUTS sampler. We also check the chain convergence for the posterior samples.

3.6 Forward modeling and analysis systematics

The first main result of this study is to present a catalog-to-cosmology pipeline and demonstrate that we can control the systematic uncertainty on A_i to be within 2% (a target motivated by our arguments in §3.2). We begin this section by laying out the algorithm that generates the maps. We then identify and catalog the systematics that significantly bias the cosmological results and provide the appropriate remedies. This section is a detailed view of the "forward model" box in Fig. 3.3.

3.6.1 Map making

Let us first consider generating a set of correlated tomographic convergence maps with Gaussian priors. To do this, we first generate independent standard normal variables (which we shall call the latent map parameters) $\tilde{\mathbf{x}}_i(\boldsymbol{\ell})$ on the Fourier grid. Next, we transform these latent parameters to the Fourier space convergence maps $\tilde{\kappa}_i(\boldsymbol{\ell})$ by (we will denote all Fourier space quantities using $\tilde{\cdot}$)

$$\tilde{\kappa}_i(\boldsymbol{\ell}) = L_{ij}(\ell)\tilde{\mathbf{x}}_j(\boldsymbol{\ell}),, \tag{3.17}$$

where $L_{ij}(\ell)$ is the Cholesky decomposition of $A_i A_j C_{ij}(\ell)$. In the trivial example with one tomographic bin and the amplitude fixed to its fiducial value ($A = 1$), this would simply mean multiplying $\mathbf{x}(\boldsymbol{\ell})$ by $\sqrt{C(\ell)}$, the RMS of the kappa map.

Now let us discuss the log-normal forward model. For the log-normal maps, we first need to compute the power spectrum of the Gaussian transfer map, this is given by

$$C_{ij}^{G} = \mathcal{H}\left[\log\left(\frac{\mathcal{H}^{-1}\left[A_i A_j C_{ij}\right]}{a_i a_j} + 1\right)\right] \tag{3.18}$$

where $\mathcal{H}[\cdot]$ is the Hankel transform.

We then generate a Gaussian transfer map just as we did in the Gaussian model, Fourier transform it into the real space, and apply the log-normal transformation in Eq. 3.6 to obtain the log-normal map.

For both the Gaussian and the log-normal model, once we obtain the κ_i, we compute the shear maps using the Kaiser-Squire relationship [113], which is most easily expressed in the Fourier space by

$$\tilde{\gamma}_1(\boldsymbol{\ell}) = \mathbb{K}_1(\tilde{\kappa}) = \frac{\ell_x^2 - \ell_y^2}{|\boldsymbol{\ell}|^2}\tilde{\kappa}(\boldsymbol{\ell}) \tag{3.19}$$

$$\tilde{\gamma}_2(\boldsymbol{\ell}) = \mathbb{K}_2(\tilde{\kappa}) = \frac{2\ell_x \ell_y}{|\boldsymbol{\ell}|^2}\tilde{\kappa}(\boldsymbol{\ell}) \tag{3.20}$$

3.6.2 Smoothing and aliasing

The process of pixelizing galaxy catalogs into shear maps (§ 3.4) introduces field-level statistical artifacts - smoothing and aliasing - in $\gamma^{\rm obs}$. Both effects introduce corrections to C that are on the same order as C itself on small scales. For exam-

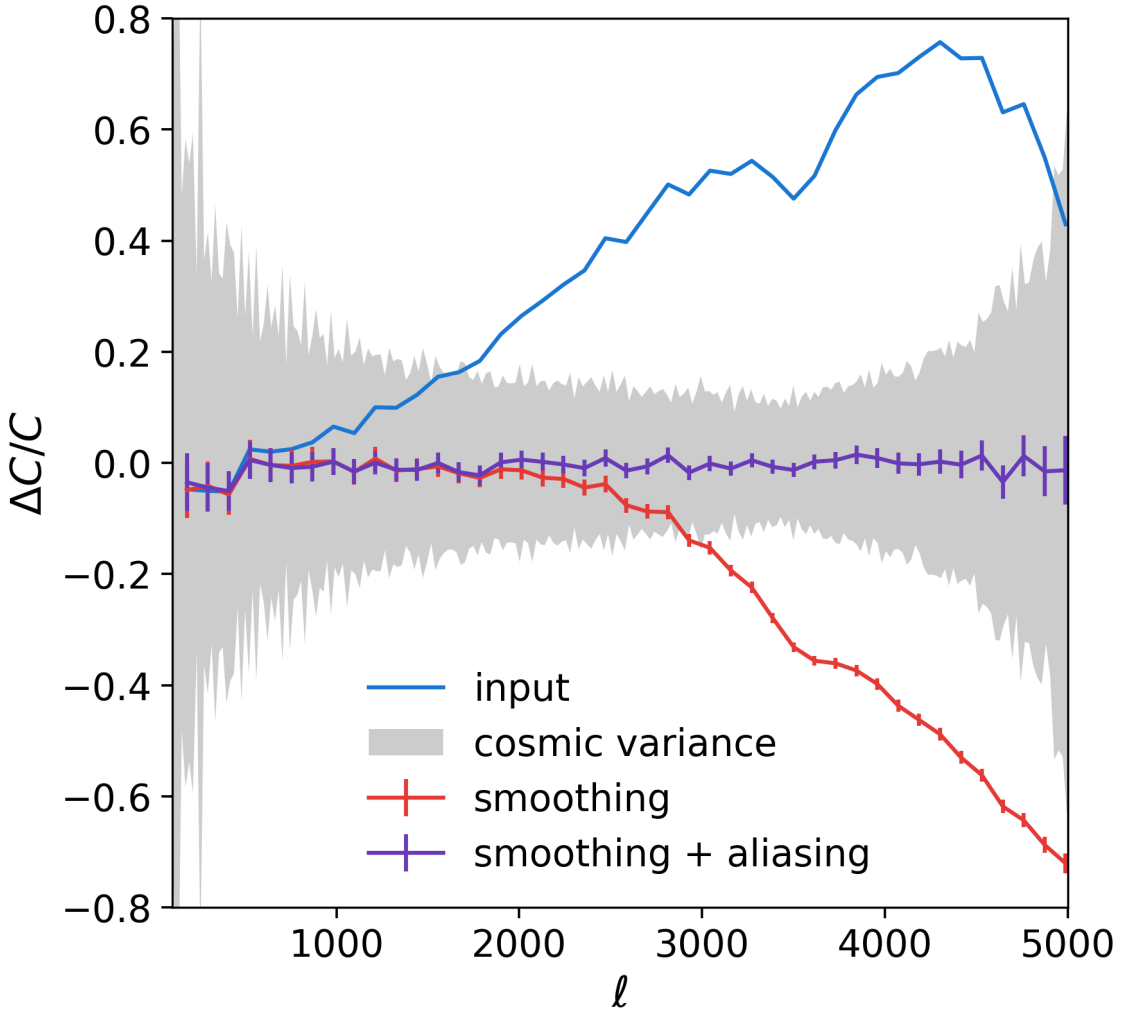

Figure 3.4: The impact of smoothing and aliasing on C is presented in terms of the fractional difference relative to $\langle C(\kappa^{\rm true})\rangle$, with the gray region representing the 1σ spread of $C(\kappa^{\rm true})$. The full-sky power spectrum is shown in blue. $\overline{C(\kappa^{\rm pix})}$, generated by a forward model that accounts for smoothing but not aliasing, is shown in red. When both effects are included in the forward model (with aliasing parameter $r = 1.6$), $\overline{C(\kappa^{\rm pix})}$ is depicted in purple. Error bars indicate the standard error of the mean. Neglecting either the smoothing or aliasing effects, or both, in the forward model results in an order-of-unity bias on C.

ple, Fig. 3.4 shows that the full-sky theory C (blue) and the distribution of $\kappa^{\rm true}$ for Gaussian mocks (gray) differ by as much as 80%. An unbiased forward model should generate $\kappa^{\rm pix}$ such that $\overline{C(\kappa^{\rm pix})} = \overline{C(\kappa^{\rm true})}$. Neglecting the pixelization effect significantly biases the cosmological results. To the best of our knowledge, this is the first precise quantification and correction of the aliasing effect in the context of field-level cosmology.

Let us start by drawing a large number of galaxies from a continuous and noiseless convergence field κ. Mathematically, the process of map-making (§ 3.4) is equivalent to first smooth κ with a pixel kernel and then sampling it at the center of each pixel (Fig. 3.5). Both effects are most easily described in the Fourier space. Smoothing is described by

$$\tilde{\kappa}^{\rm smooth} = {\rm sinc}\left(\frac{\ell_x\Delta}{2\pi}\right){\rm sinc}\left(\frac{\ell_y\Delta}{2\pi}\right)\ \tilde{\kappa},, \tag{3.21}$$

Meanwhile, the innocent-looking sampling procedure not only discretizes the Fourier space but also aliases κ. The net effect increases the small-scale (near the Nyquist frequency of the observed map, $\ell_{\rm N} = \pi/\Delta$) power of the pixelized map by an order of unity and breaks the statistical isotropy of κ. More precisely, pixelizing the continuous field κ means [114, 115]

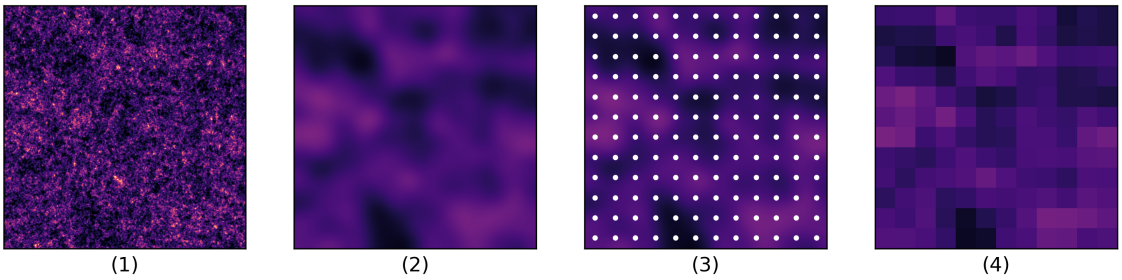

Figure 3.5: Illustration of the mathematical operations implied by the map-making process (§ 3.4). Plot (1) shows a continuous field. During map-making, the field is first smoothed with a pixel kernel (2). The smoothed field is then sampled at the pixel centers, indicated by white dots in (3). The values at these pixel centers constitute the pixelized map seen in (4). Both the smoothing and sampling processes introduce significant statistical artifacts into the pixelized maps, which must be accounted for in the forward model.

$$\tilde{\kappa}^{\rm alias}(\boldsymbol{\ell}) = (\tilde{\rm III} * \tilde{\kappa})(\boldsymbol{\ell}) + \underbrace{\sum_{(k_1,k_2)\neq(0,0)} (\tilde{\rm III} * \tilde{\kappa})(\ell_x - k_1\ell_{\rm N}, \ell_y - k_2\ell_{\rm N}),,}_{\text{aliases}} \tag{3.22}$$

$$C^{\rm alias}(\boldsymbol{\ell}) = C(\boldsymbol{\ell}) + \underbrace{\sum_{(k_1,k_2)\neq(0,0)} C_{ij}(\ell_x - k_1\ell_{\rm N}, \ell_y - k_2\ell_{\rm N}),.}_{\text{aliases}} \tag{3.23}$$

We define $*$ as the (Fourier space) convolution and the Kronecker comb function as

$$\text{III}(\boldsymbol{\theta}) = \sum_{n_x,n_y\in\mathbb{Z}} \delta_{\rm K}\left(\theta_x - n_x\Delta\right)\delta_{\rm K}\left(\theta_y - n_y\Delta\right),, \tag{3.24}$$

where $\delta_{\rm K}(x) = 1$ when $x = 0$ and 0 otherwise.

In short, Eq. 3.22 tells us that each mode of $\tilde{\kappa}^{\rm alias}(\boldsymbol{\ell})$ is a superposition of $\tilde{\kappa}(\boldsymbol{\ell})$ and its "aliases" at higher frequencies. Eq. 3.23 tells us that the power spectrum of the pixelized map acquires a two-dimensional $\boldsymbol{\ell}$ dependence and is no longer isotropic.

Furthermore, $C^{\rm alias}(\boldsymbol{\ell})$ is also the superposition of $C(\boldsymbol{\ell})$ and its aliases. One immediate consequence is that, around the Nyquist frequency, $C^{\rm alias}$ is at least 100% higher than C. Although Eq. 3.22 suggests that accounting for infinitely many aliases at increasingly larger ℓ's is necessary to fully capture aliasing, in practice, aliases above $\sqrt{2}r\ell_{\rm N}$ with $r = 1.6$ are sufficiently suppressed by the smoothing kernel.

We implement Eqs. 3.21 and 3.22 in `Miko` to model the pixelization effect. The results are demonstrated in Fig. 3.4 for the Gaussian model. If we only include smoothing (red) in the forward model, the $\overline{C(\kappa^{\rm pix})}$ is two times lower near $\ell_{\rm N}$, as expected. When both smoothing and aliasing are included (purple), the model generates $\kappa^{\rm pix}$'s that align with $\kappa^{\rm true}$ on the two-point level across all scales. This behavior is consistently observed across all the cross-power spectra and is also confirmed for the log-normal model.

We now test the bias introduced by the pixelization effect on the cosmological parameters A_i. We setup three different Gaussian models. Each model includes the smoothing correction but differs in the aliasing parameter r, set at 1 (no aliasing correction), 1.3, and 1.6 respectively. We run each model on 80 HSC-like experiments, using Gaussian mocks. The results are presented in Fig. 3.6. Neglecting aliasing in the forward model biases A_i higher. The direction of the bias is expected since aliasing amplifies the small-scale power in $\gamma^{\rm obs}$ which can only be compensated by increasing A_i. Neglecting aliasing leads to a 2.5% absolute bias on the cosmological parameters exceeding our error budget. However, setting $r = 1.6$ is generally sufficient to model the aliasing effect. We will use $r \geq 1.6$ for all the following analyses.

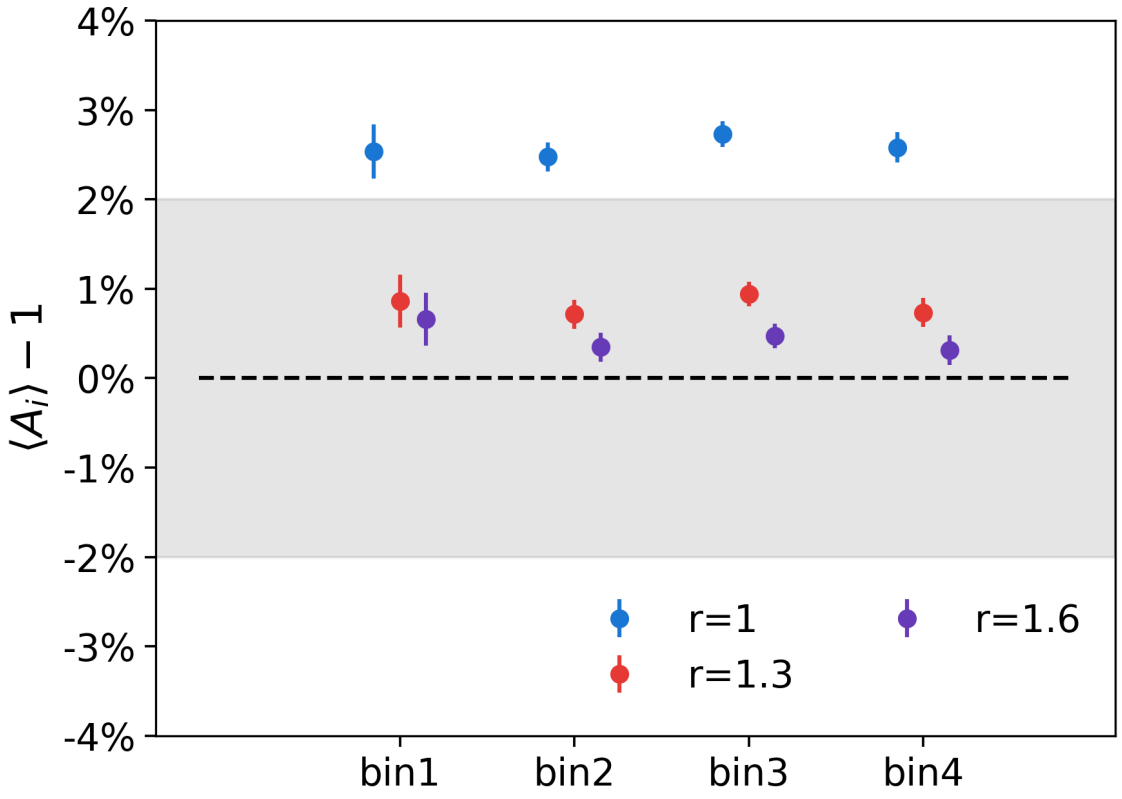

Figure 3.6: The absolute bias of A_i measured by applying the Gaussian model to Gaussian mocks. Each color corresponds to a Gaussian model with a different aliasing parameter r. Each data point represents the average inferred A_i over 80 independent HSC-like experiments, with the error bars denoting the standard error of the mean. The experiment is set with an 80% margin (see § 3.6.3 and Fig. 3.8). The gray region indicates the 2% systematics control target.

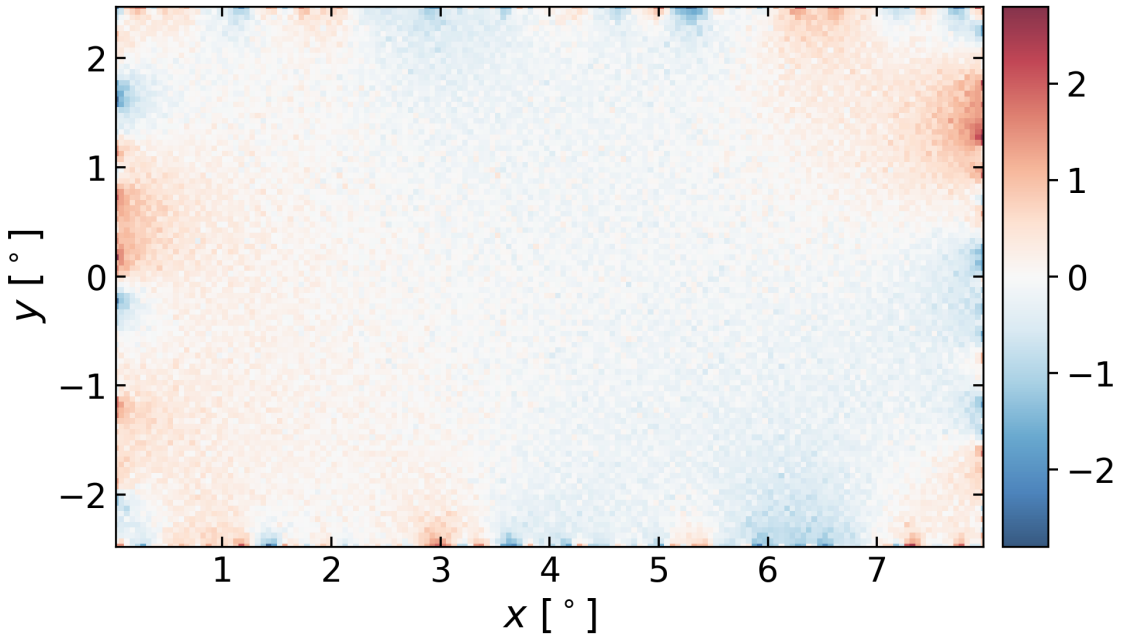

Figure 3.7: The Kaiser-Squire reconstruction error of a (Gaussian) shear field assuming periodic boundary condition. We define the reconstruction error as $\epsilon = (\mathbb{K}_1\kappa - \gamma_1)/\Delta_\gamma$, where $\mathbb{K}_1$ is the Kaiser-Squire operator that transforms κ to γ_1 assuming a nonperiodic boundary at spatial infinity. Δ_γ is the standard deviation of the shear field. We observe the most error along the boundaries.

3.6.3 Boundary conditions

We use the Kaiser-Squire relationship Eqs. 3.19 and 3.20 to transform κ into γ in the forward model. The Kaiser-Squire relationship is exact on an infinite plane. However, our forward model is defined on a finite grid which has a periodic boundary condition inherited from the fast Fourier transform. This mismatch introduces field-level errors around the edges of the maps as shown in Fig. 3.7. We can alleviate this bias by generating maps larger than the survey footprint and defining the likelihood on the inner regions. More specifically, we test how different amounts of margins, measured

by the fractional length of the short side of the map, affect the final cosmological parameter constraints. The result is shown in Fig. 3.8. We find that the boundary condition mismatch drives A_i higher. Furthermore, a 80% margin is sufficient to reduce the bias below 0.5%. We will use this choice of margin for all the following analyses.

3.6.4 Fourier mode coupling

As pointed out by Xavier *et al.* [104] and Tessore *et al.* [116], the log-normal model transformation is not local in Fourier space and strongly couples the small-scale modes. A finite Fourier grid implies a strict band-limit and therefore changes the expected behavior of the transformation. To faithfully simulate a log-normal map up to $\ell_{\rm max}$, we need to generate a Gaussian transfer map up to $\ell^G > \ell_{\rm max}$. For example, $\ell^G > 4\ell_{\rm max}$ is required for percent-level accuracy when simulating full-sky log-normal maps using `FLASK` (as is done for our log-normal mocks § 3.4). The iterative correction method proposed by Tessore *et al.* [116] can, in theory, produce logarithmic normal maps with accurate power spectra with $\ell^G \approx \ell_{\rm max}$. We, however, did not achieve the same success when applying the same idea to the flat sky case, potentially due to the boundary condition and the nonisotropy of the Fourier grid.

Nevertheless, we found an empirical solution to reduce this error. In Fig. 3.9, we plot the scale-dependent error of the log-normal $C(\kappa)$ relative to the input C as a function of ℓ^G. If we set ℓ^G to the maximum aliasing scale $\sqrt{2}r\ell_{\rm N}$ (blue), $C(\kappa)$ starts to deviate from the input beyond $\ell_{\rm N}$ (vertical line). To obtain the correct $C(\kappa)$ beyond $\ell_{\rm N}$, we find it useful to force $\kappa^{\rm G}$ to be isotropic by removing powers

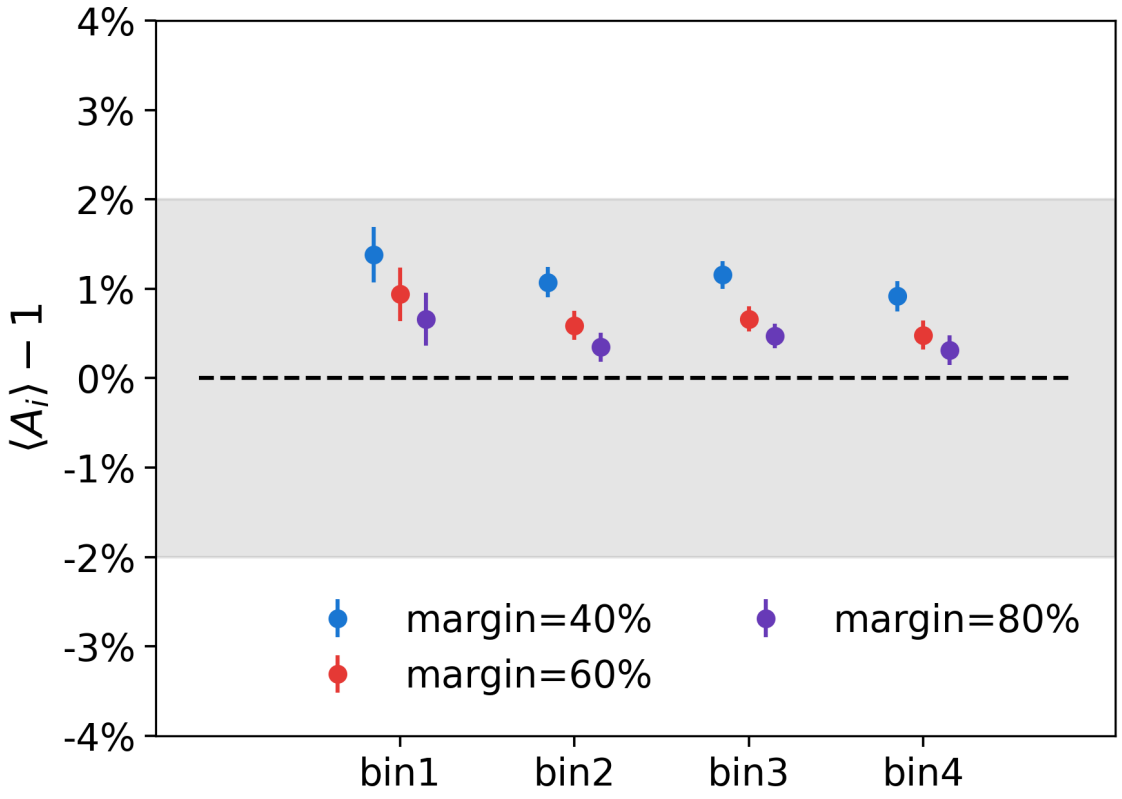

Figure 3.8: The absolute bias of A_i measured by applying the Gaussian model to Gaussian mocks. Each color corresponds to a Gaussian model with a different margin size (as a fraction of the length of the short side of the patch). Each data point represents the average inferred A_i over 80 independent HSC-like experiments, with the error bars denoting the standard error of the mean. The experiment is set with $r = 1.6$ (see § 3.6.2 and Fig. 3.6). The purple points corresponds to the same purple points in Fig. 3.6. The gray region indicates the 2% systematics control target.

beyond $r\ell_{\rm N}$ (red). If we impose an even stricter scale-cut, the log-normal maps exhibit lower band power below $\ell_{\rm N}$. Regardless of ℓ^G, the log-normal maps still have a deficit of large-scale power. This appears to be a consequence of the small survey footprint or the periodic boundary effect, and this error is reduced when generating larger maps. For the log-normal model, since the Fourier mode coupling issues that we have discussed so far interact together, it is difficult to quantify each error separately. However, looking ahead, we will show that using $\ell^G = r\ell_{\rm N}$ with $r = 1.9$ enables us to control the systematic bias for the log-normal model to 2%.

3.6.5 Number density-induced shape noise

Finally, we observe another novel source of error that impacts FLI—the finite galaxy density can induce a shape noise-like error on the pixelized maps even when no shape noise is introduced. When we average the galaxy shapes within a pixel, the distribution of random galaxy positions samples the inhomogeneous shear field, leading to statistical fluctuation of the pixel value. Formally, for a pixel p centered at $\boldsymbol{\theta}_p$, we expect the galaxy's true shape to differ from $\gamma(\boldsymbol{\theta}_p)$ by

$$\Delta\gamma \approx \frac{\partial\gamma}{\partial\theta}(\boldsymbol{\theta}_p)\Delta\theta = \ell\gamma\Delta\theta, . \tag{3.25}$$

This term behaves like a white noise that only appears in the autopower spectra of $\gamma^{\rm pix}$. Therefore, we call this the number density-induced shape noise, or density shot noise in short. The impact of density shot noise is characterized in Fig. 3.10, where we see that it impacts the small scale autopower spectrum by about 20% near the

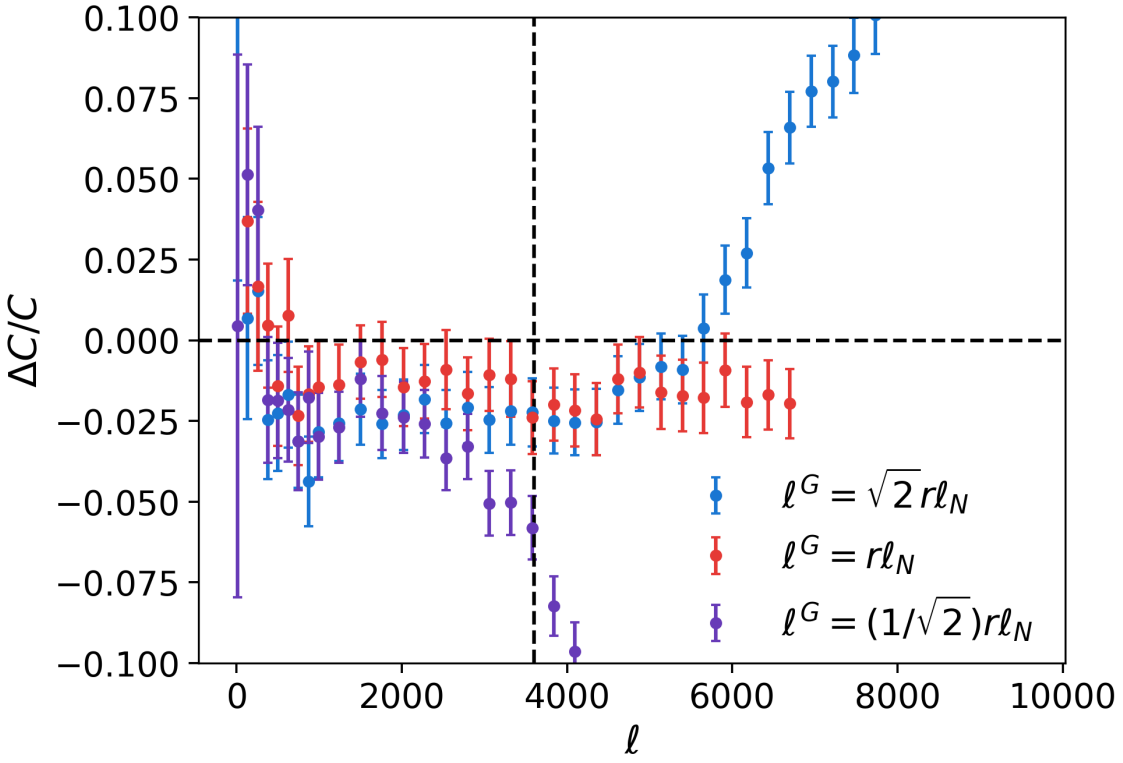

Figure 3.9: The impact of the Gaussian transfer map scale-cut, ℓ^G, on the recovered C is presented in terms of the fractional difference relative to the input C. The error bars represent the standard error of the mean. The vertical line denotes the Nyquist frequency, $\ell_{\rm N}$, of $\gamma^{\rm obs}$. Setting $\ell^G = r\ell_{\rm N}$ (red) recovers the input C with the highest accuracy. The observed residual additive bias is related to the limited sky coverage of the patch.

pixel resolution scales. We did not observe this noise in the crosspower spectra.

To correct for this systematics, we first measure the power spectrum of this density-induced shape noise term ($C_{ii}^{\rm dsn}$) directly from mock catalogs for each redshift bin i (as shown in Fig. 3.10). Then, when we generate convergence maps following Eqs. 3.17 (for the Gaussian model) and 3.18 (for the log-normal model), we can simply change the cosmological power spectrum C_{ii} via

$$C_{ii} \to C_{ii} + C_{ii}^{\rm dsn}, . \tag{3.26}$$

In this way, the shear maps ($\gamma_i^{\rm pix}$) generated by the forward model automatically include the density shot noise. The expression $(\gamma_i^{\rm obs} - \gamma_i^{\rm pix}(\mathbf{c}, \mathbf{x}_i))^2/2N$ in Eq. 3.16 still accurately describes the regular shape noise model and does not require additional modification.

3.7 Model misspecification

We have demonstrated that by accounting for various analysis systematics —including aliasing, boundary effects, mode coupling, and density-induced shape noise —we can obtain unbiased cosmological constraints from observed pixelized shear maps. A remaining modeling uncertainty is the field-level prior on the convergence field. It is anticipated that this uncertainty could potentially induce an absolute bias on the inferred parameters as well as affect the uncertainty quantification. In this section, we test both the Gaussian and the log-normal models using Gaussian, log-normal, and Takahashi mocks.

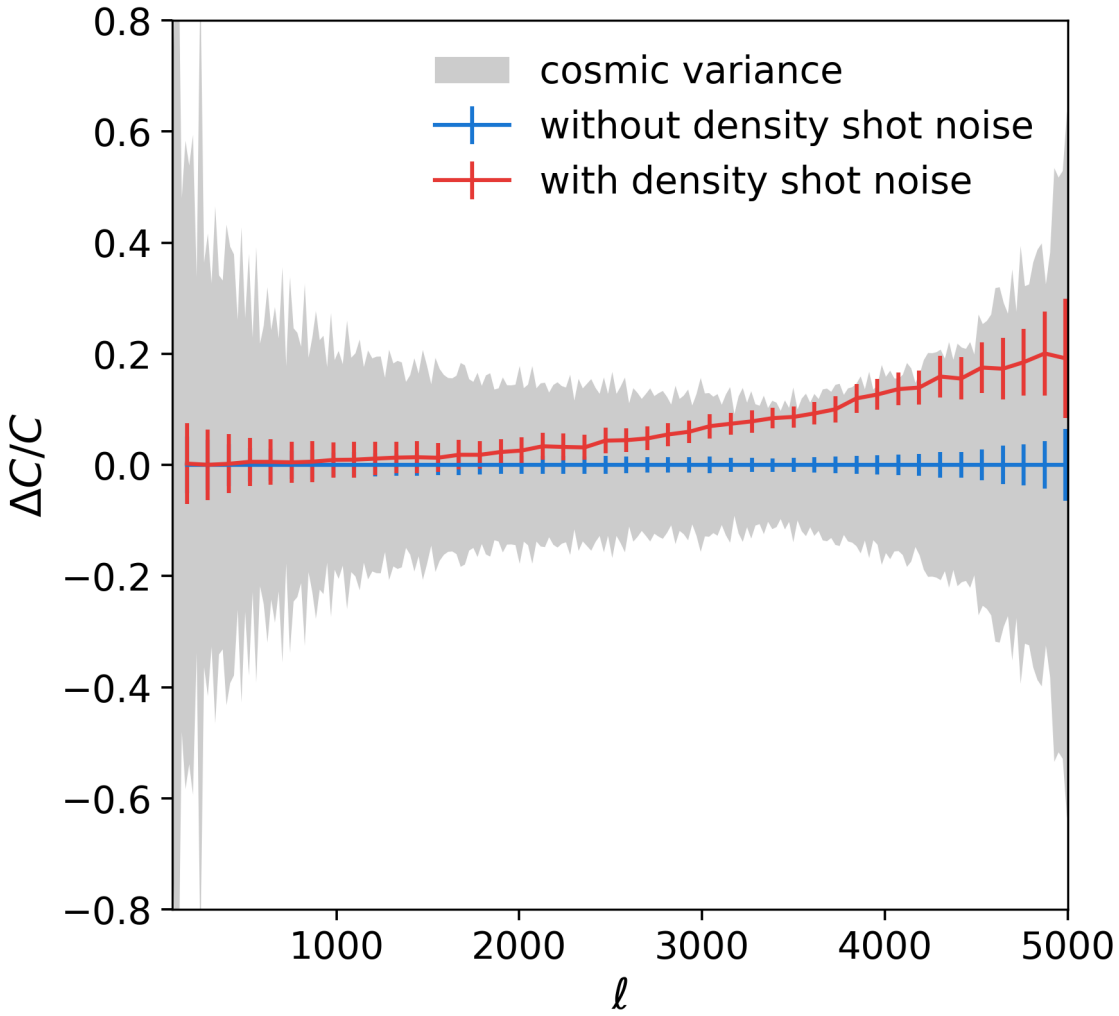

Figure 3.10: The impact of number density-induced shape noise on C is presented in terms of the fractional difference relative to $\langle C(\kappa^{\rm true})\rangle$ where the number density-induced shape noise is removed (achieved by creating maps using ρ_i and $\gamma^2_{\rm rms}$ at ten times the fiducial value). The high density case is represented in blue, and its 1σ spread is shown in gray (as in Fig. 3.4). $\overline{C(\kappa^{\rm true})}$ calculated with fiducial values for ρ_i and $\gamma_{\rm rms}$ is shown in red. The error bars denote the standard error of the mean.

As we have shown above, a significant difference exists between the realistic observed shear maps and the maps generated by the forward model (which we shall call pipeline mocks). Previous works [76, 87] have demonstrated that it is possible to use the forward model to analyze the pipeline mocks and correctly recover the cosmological parameters. Here, we follow Fiedorowicz *et al.* [47] and go one step further toward realism by using `Miko` to analyze the raw mocks themselves. A further feature of this analysis is that, to the best of our knowledge, this is the first study of its kind to extract cosmological parameters (here the A_i) using realistic mocks and taking into account the range of analysis systematics that must be considered for real data analysis.

3.7.1 Impact on cosmological parameters

The Gaussian model

First, we run the Gaussian model on all three sets of mocks. For each model-data pair, we perform 80 HSC-like experiments on independent patches, utilizing analysis systematics models recommended by § 3.6. For each experiment, we obtain the joint posterior distribution of A_i along with the map parameters. For instance, the constraint on A_i for a single Takahashi patch is depicted as the red contour in Fig. 3.11. The A_i's are uncorrelated in the Gaussian model, consistent with the findings in Zhou and Dodelson [24]. There is considerable cosmic variance within a $40\,\mathrm{deg}^2$ patch. To determine whether the pipeline has an absolute bias on A_i, we compute $\overline{A_i} - 1$ over independent experiments and present the results in the top

panel of Fig. 3.12. The top part of the top panel shows that the Gaussian model does not exhibit absolute bias for all simulated datasets, regardless of whether the data is simulated with a Gaussian model or not.

To assess the impact of model misspecification on our uncertainty estimates, let us consider the one-dimensional marginal posterior distribution of a particular A_i. Ideally, $x\%$ (the observed probability) of all the experiments should find $A_i = 1$ contained within the $x\%$ (the expected probability) credible interval. The observed probability as a function of the expected probability is called the P-P plot. It is shown in the lower part of the top panel of Fig. 3.12, after adjusting for mean absolute bias to concentrate on the error bars. Perfect uncertainty prediction would align the P-P curves with the line $y = x$. This alignment occurs only without model misspecification. For instance, applying the Gaussian model to log-normal or Takahashi mocks results in overly optimistic uncertainty estimates for all A_i values, particularly for log-normal mocks where the truth is included within the one standard deviation interval less than 20% of the time. This is consistent with the observation of Boruah *et al.* [76], although the extent to which the Gaussian model underestimates the uncertainty is much more severe in our case. The bottom line is that assuming a Gaussian prior leads to an unbiased estimate of the cosmological parameters but a *biased* estimate of the errors on those parameters.

The "unreasonable" effectiveness of the Gaussian priors

Why does the Gaussian map prior not induce any absolute bias on A_i when it is applied to clearly non-Gaussian fields? The answer lies in the PDF of $\tilde{\kappa}$ in Fourier

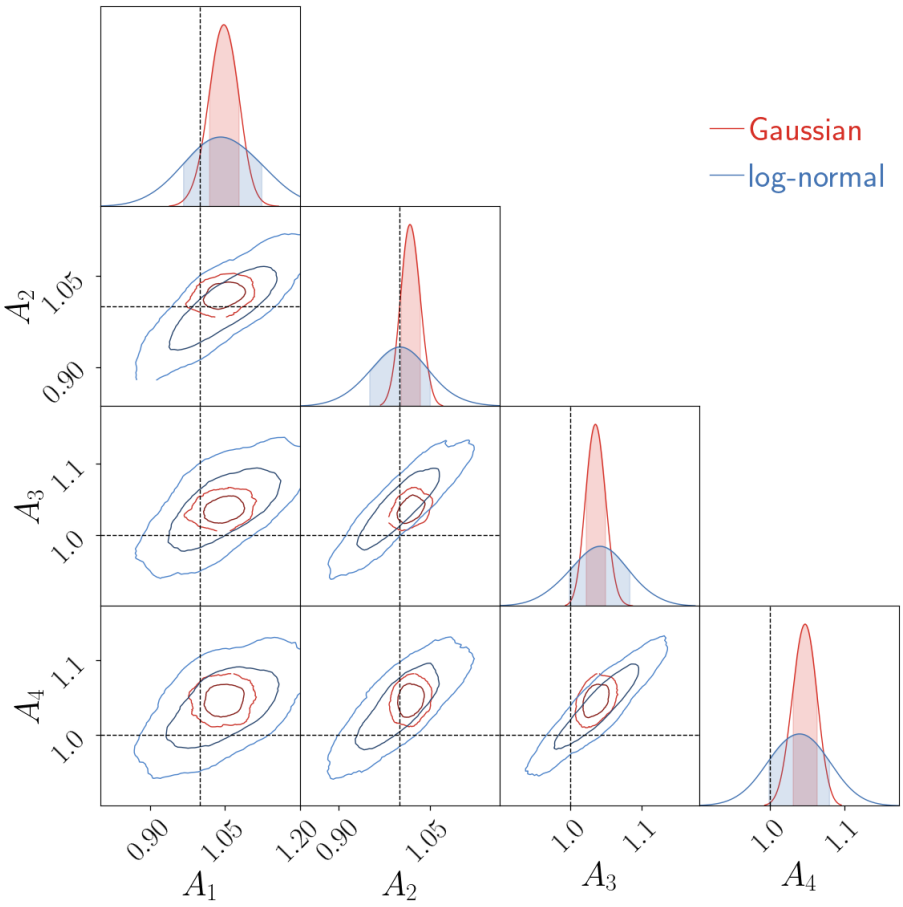

Figure 3.11: The joint posterior distribution of A_i inferred from a 40 deg^2 Takahashi patch using the Gaussian (red) and the log-normal (blue) models. The log-normal model employs an informative prior on a_i that is jointly sampled (not shown here) with A_i. Compared to the Gaussian model, the log-normal model exhibits larger uncertainties and a stronger correlation between the A_i's. The contours show the 1σ and 2σ level sets.

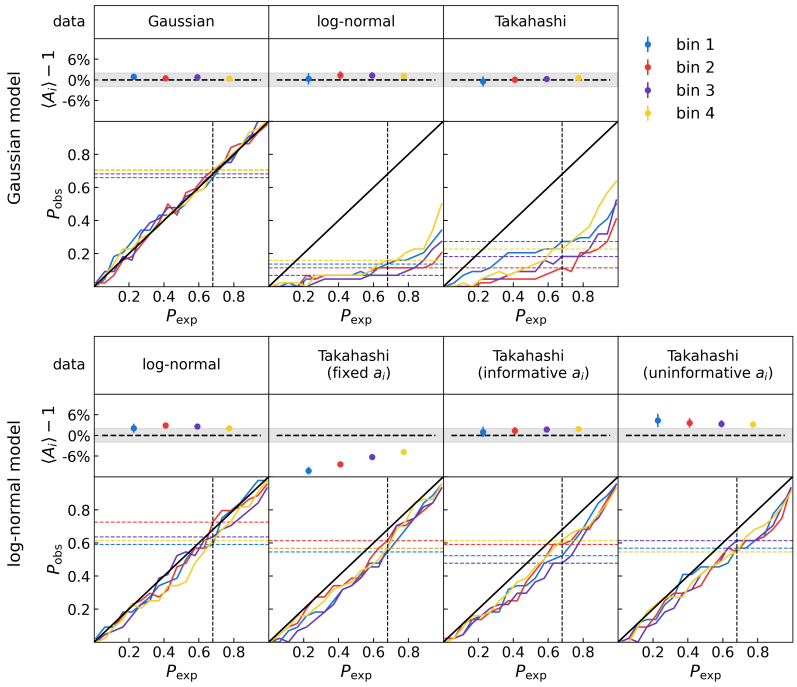

Figure 3.12: The top panel displays the results of the Gaussian model applied to Gaussian, log-normal, and Takahashi mocks. Within this panel, the first row indicates the absolute bias of A_i, averaged over 80 independent HSC-like experiments. The gray area denotes the systematic error control target of 2%. The second row presents the P-P plot for A_i (with the mean bias subtracted), where the x-axis denotes the expected probability ($P_{\rm exp}$) that a credible interval contains the true A_i, and the y-axis denotes the observed probability ($P_{\rm obs}$). The vertical dashed line indicates $P_{\rm obs} = 68\%$, and the colored horizontal lines show where the P-P curves intersect with this line. The Gaussian model demonstrates no absolute bias irrespective of model misspecification but exhibits overconfidence in the error bars when applied to realistic data. The bottom panel illustrates the absolute bias and the P-P plot for A_i using the log-normal model, applied to log-normal and Takahashi mocks. For the Takahashi mocks, three scenarios are tested with the log-normal model concerning the shift parameters: 1) fixed at values measured from noiseless true mocks (fixed a_i), 2) allowed to vary within a narrow and informative range, and 3) allowed to vary within a broad and uninformative range. The log-normal model generally yields accurate uncertainties, yet the absolute bias is highly sensitive to the choice of a_i.

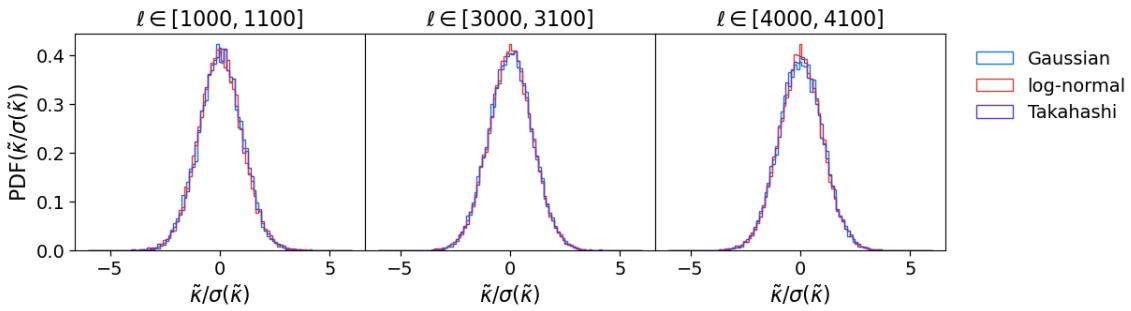

Figure 3.13: Standardized one-point PDF of $\tilde{\kappa}(\ell)$ for Gaussian, log-normal, and Takahashi mocks at three distinct ℓ values. Across all mocks and ℓ values, the distribution of $\tilde{\kappa}(\ell)$ remains consistently Gaussian, which elucidates the Gaussian model's ability to infer cosmological parameters without absolute bias, notwithstanding model misspecification.

space. In Fig. 3.13, we plot the PDFs of $\tilde{\kappa}^{\rm true}$ for Gaussian, log-normal, and Takahashi mocks at three different ℓ values. Interestingly, for all mocks and across all ranges of ℓ, the PDF of $\tilde{\kappa}^{\rm true}$ is consistent with a Gaussian distribution. Similar phenomena have previously been observed in three-dimensional simulations, for example, by Matsubara [117] and Qin *et al.* [118]. Motivated by these observations, we can express the PDFs of non-Gaussian ($\text{prior}^{\rm NG}$) and Gaussian ($\text{prior}^{\rm G}$) convergence fields as follows (ignoring aliasing and other analysis systematics for simplicity)

$$\text{prior}^{\rm NG}(\tilde{\kappa}) = \text{prior}^{\rm G}(\tilde{\kappa})\left[1+\delta(\tilde{\kappa})\right],. \tag{3.27}$$

In the absence of model misspecification, the likelihood analyses should yield unbiased $A_i^{\rm NG}$, where

$$\begin{aligned}
&\langle A_i^{\rm NG}\rangle \\
&= \int d\gamma^{\rm obs} \int d\tilde{\kappa}\ A_i\ \mathcal{P}(\gamma^{\rm obs}|\tilde{\kappa}, A_i)\ {\rm prior}^{\rm NG}(\tilde{\kappa}) && (3.28)\\
&= \int d\gamma^{\rm obs} \int d\tilde{\kappa}\ A_i\ \mathcal{P}(\gamma^{\rm obs}|\tilde{\kappa}, A_i)\ {\rm prior}^{\rm G}(\tilde{\kappa})\,[1+\delta(\tilde{\kappa})] && (3.29)\\
&= \langle A_i^{\rm G}\rangle + {\rm correction}\,[\delta] && (3.30)
\end{aligned}$$

Therefore, as long as the perturbation δ is small, $\langle A_i^{\rm G}\rangle = \langle A_i^{\rm NG}\rangle$ is unbiased.

The second question pertains to why the Gaussian model fails to predict the correct uncertainty of A_i. To address this, we first examine the correlation between $C(\ell_1)$ and $C(\ell_2)$ across the Gaussian model, the log-normal model, and the Takahashi mocks, as illustrated in Fig. 3.14. As expected, the Gaussian model assumes there is no correlation between the power of two different ℓ modes. Nonetheless, correlations are evident in both the log-normal simulations and the realistic Takahashi mocks. Consequently, the Gaussian model presupposes that each ℓ mode contributes independently, while in actuality, the different ℓ modes are correlated and thus provide less information than assumed. As a result, the Gaussian model reports a smaller error bar on the inferred power spectrum amplitude A_i, an error analogous to assuming a larger $f_{\rm sky}$ than what is true in reality. Another perspective on this issue is provided by examining the cosmic variance predicted by the Gaussian, log-normal, and Takahashi maps, as shown in Fig. 3.15. The Gaussian model predicts smaller cosmic variance than the other two models across all scales, which then translates to

the narrower uncertainty bounds on the estimated power spectrum amplitude A_i.

The log-normal model

The log-normal model depends on both A_i and the shift parameter a_i. We first run the log-normal model on the 44 log-normal mocks fixing a_i at the truth. The pipeline recovers the correct A_i with an absolute bias within 2%[3]). We also report excellent uncertainty quantification. This result is shown in the bottom panel of Fig. 3.12 (column 1).

Model misspecification also arises when applying the log-normal model to the Takahashi mocks. Specifically, since the Takahashi mocks do not adhere to a strict log-normal distribution, as shown in Fig. 3.1, a definitive ground truth for a_i is absent. Previous studies have approached a_i in various ways: for instance, Alsing *et al.* [87] treated a_i as fixed parameters, whereas Boruah *et al.* [76] and Fiedorowicz *et al.* [47] employed perturbation theory for approximation. As we will show, the modeling of a_i is arguably the primary source of bias with the log-normal model and warrants careful consideration.

In our first approach, we compute $a_i^{\rm fit}$ by fitting the one-point PDF to the noiseless Takahashi mocks using Eq. 3.7. We then perform inference on the Takahashi mocks assuming $a_i = a_i^{\rm fit}$. This method approximates a perfect perturbation theory prediction for a_i. The result is presented in the bottom panel of Fig. 3.12 (column 2). Surprisingly, even when using the optimally fitted $a_i^{\rm fit}$, the log-normal model

3. The slight positive bias is due to the deficit of large-scale powers discussed in § 3.6.4. This error can be alleviated by using larger patches.

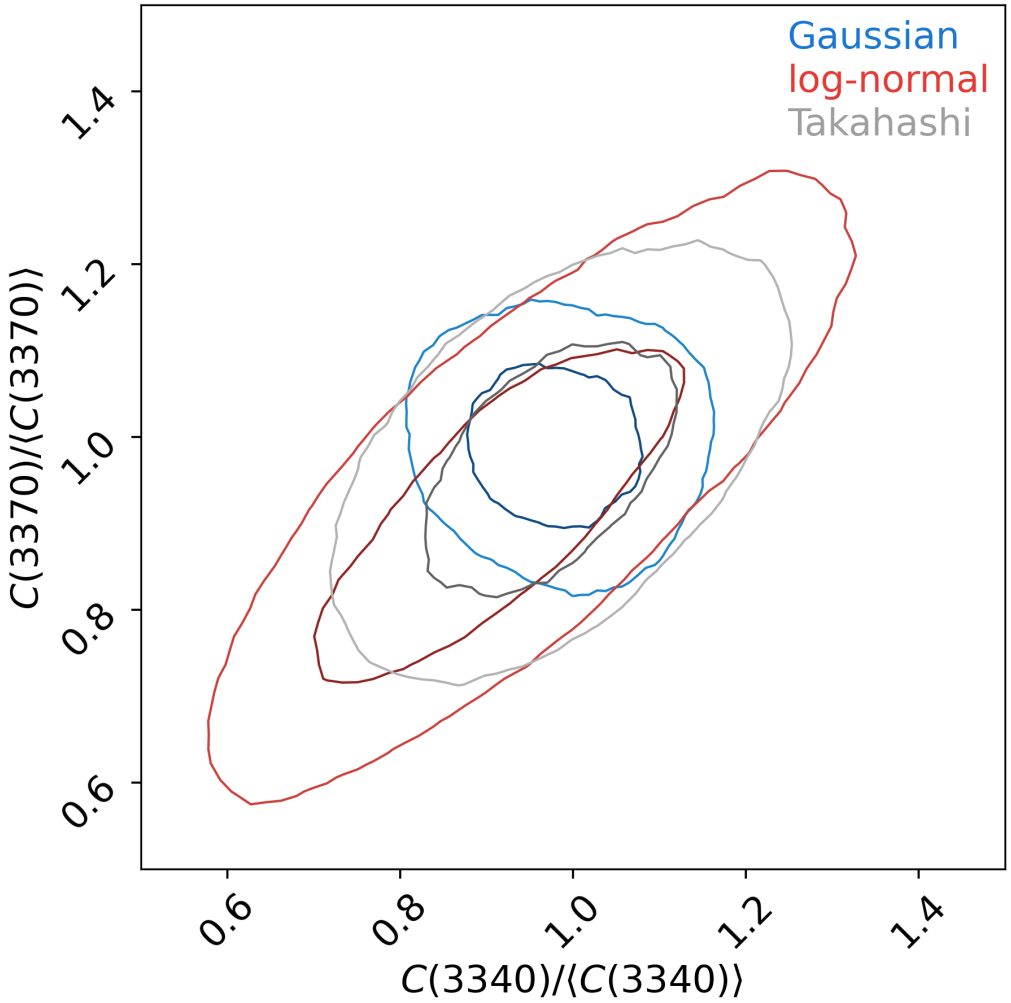

Figure 3.14: The correlation between $C(\ell_1)/\langle C(\ell_1)\rangle$ and $C(\ell_2)/\langle C(\ell_2)\rangle$ for Gaussian (blue), log-normal (red), and Takahashi (gray) mocks. In this example, we have chosen $\ell_1 = 3340$ and $\ell_2 = 3370$, and C to be the autopower spectrum of the first redshift bin. The contours show the 0.3 and 0.8σ level sets. The Gaussian model assumes no mode correlations. The log-normal model assumes more mode correlation than the realistic mocks.

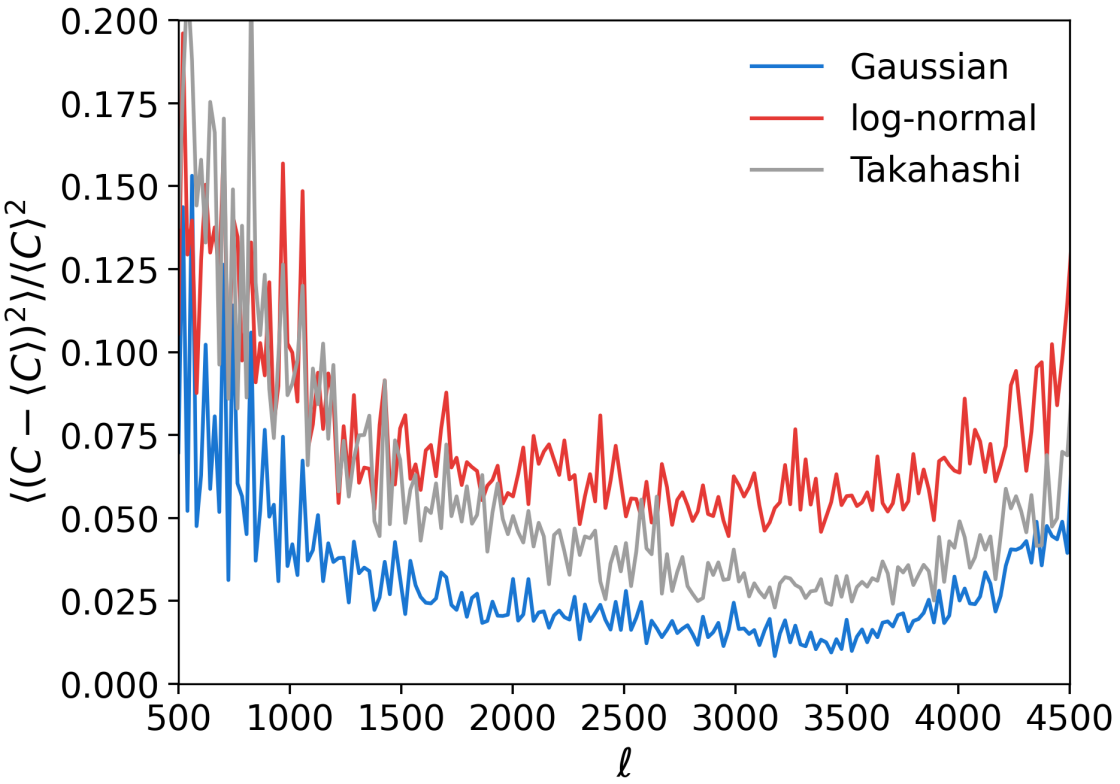

Figure 3.15: Normalized variance of the power spectrum for Gaussian (blue), log-normal (red), and Takahashi (gray) mocks, with C representing the autopower spectrum of the first redshift bin. Notably, the Gaussian mocks exhibit smaller cosmic variance relative to the Takahashi mocks, whereas the log-normal mocks demonstrate greater cosmic variance in comparison.

infers A_i with a significant negative bias. This bias is most pronounced, reaching up to 10%, in the lowest redshift bins where the convergence field exhibits the highest non-Gaussianity. Consequently, we deduce that merely employing the log-normal shift parameter that most closely fits the PDF of the actual convergence field is not only insufficient but also incorrect for acquiring unbiased constraints on cosmological parameters.

In our second approach, we adopt an agnostic stance on a_i, treating it as a variable during sampling. Observing that the likelihood analysis typically favors $a_i > a_i^{\rm fit}$, we implement a flat prior on a_i in the range $\left[0.9a_i^{\rm fit}, 1.3a_i^{\rm fit}\right]$. The joint posterior of A_i, as depicted in Fig. 3.11, reveals that the uncertainties reported by the log-normal model are larger—yet appear more realistic—than those by the Gaussian model, with A_i exhibiting positive correlation. Furthermore, the joint posterior of A_i and a_i (Fig. 3.16) indicates a positive correlation between them, with the maximum likelihood estimate for a_i substantially exceeding $a_i^{\rm fit}$. This is consistent with the previous observation that fixing a_i to $a_i^{\rm fit}$ induces a negative bias on A_i. Nonetheless, Fig. 3.16 also illustrates that the prior range for a_i truncates the posterior distribution, hence we refer to this as the informative prior model. Through 44 independent experiments, we find that imposing an informative prior on a_i can significantly reduce the absolute bias on A_i to within 2% while maintaining accurate uncertainty estimation for A_i. The P-P curves in the bottom panel of Fig. 3.12 quantify the extent to which the uncertainty estimation is accurate when the log-normal prior is used.

Lastly, we impose a uniform and uninformative prior on a_i over the positive reals.

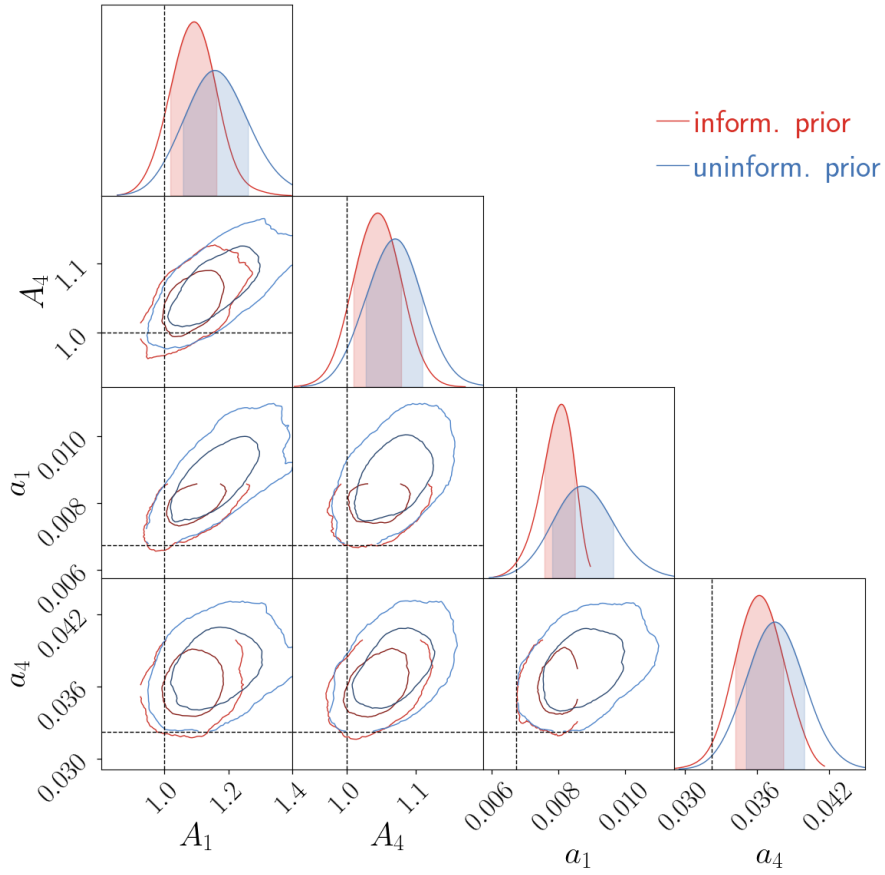

Figure 3.16: The joint posterior distribution of A_i and a_i, as inferred from a Takahashi patch using the log-normal model. We compare an informative prior (red) and an uninformative prior (blue) on a_i. For simplicity, only the first and the last redshift bins are shown. The dashed lines across A_i represent $A_i = 1$, and the dashed lines across a_i represent $a_i^{\rm fit}$—the best-fit values from noiseless data. Three important conclusions are 1) A_i and a_i are positively correlated, 2) the log-normal model prefers a_i much larger than $a_i^{\rm fit}$, and 3) the informative prior truncates the distribution of a_i. Therefore, when the prior is relaxed, A_i shifts to higher values. The contours show the 1σ and 2σ level sets.

We explicitly checks that the prior does not truncate the posterior distribution. An example is shown in Fig. 3.16. As the positive correlation between a_i and A_i suggest, this model creates a positive absolute bias on A_i (bottom panel of Fig. 3.12, column 4). The bias is most severe in the lowest redshift bin at the 4% level.

To summarize, the three choices for a_i's prior suggest that, in order to obtain unbiased cosmological constraints on real data using the log-normal model, we can neither use a_i directly measured from the noiseless mocks nor treat it as a variable that is completely free. Otherwise, although we obtain good uncertainty quantification, we risk significant bias in cosmological parameters. The existence of the absolute bias is not too surprising given that the log-normal also misspecifies the field-level priors.

3.7.2 Map reconstructions

`Miko` also provides the joint posterior distribution of all the map pixels—$(\kappa_i^{\rm pix})_s$ (the bottom diamond in the flowchart Fig. 3.3), where s denotes the sample index. For simplicity, let us focus specifically on one experiment on the Takahashi mock, and on the reconstruction of the first redshift bin convergence field. The results for both the Gaussian and log-normal (informative prior) models are summarized in Fig. 3.17.

Let us start with the first row. The first and second columns depict the observed noisy shear field, $\gamma^{\rm obs}$, and the true convergence, $\kappa^{\rm true}$, respectively. The final three columns pertain exclusively to the log-normal model, illustrating the mean posterior map $\overline{\kappa^{\rm pix}}$, the standard deviation map σ, and the signal-to-noise ratio map (S/N). For these last three maps, we zoom in on a $1.5\,\deg \times 1.5\,\deg$ patch in the inset plots

below. In these insets, we also compare the log-normal results with those of the Gaussian model.

Theoretically, we do not expect $\overline{\kappa^{\rm pix}}$ to match $\kappa^{\rm pix}$ exactly. Instead, $\overline{\kappa^{\rm pix}}$ is *roughly* analogous to a Wiener-filtered version of $\kappa^{\rm pix}$. This filtering smooths the map on small scales where shot noise dominates (see Zhou and Dodelson [24] for a detailed discussion on the properties of the posterior maps and the maps' power spectra). Both the Gaussian and the log-normal model recover the large-scale features of the true convergence maps. The log-normal model, however, resolves the density peaks much better (insets in the second row). The adjacent histogram plot further supports this claim - the log-normal model's PDF($\overline{\kappa^{\rm pix}}$) agrees with PDF($\kappa^{\rm true}$) up to the highest κ values. In contrast, the Gaussian model's PDF($\overline{\kappa^{\rm pix}}$) decays much faster. On the other hand, the log-normal model's PDF($\overline{\kappa^{\rm pix}}$) shows a hard cutoff in the low κ region. The Gaussian model captures the voids much better in comparison. This aligns with the findings of Fiedorowicz *et al.* [47], who reported that the log-normal model is less successful in reconstructing the correct void count distribution at higher resolutions.

The log-normal and the Gaussian model also differ in their prediction of the pixel-level uncertainty. (We show the 1σ uncertainty for each pixel in the fourth column and the comparison between the log-normal model and the Gaussian model in the inset plots below.) The variance predicted by the log-normal model shows a positive correlation with the absolute pixel value, whereas the Gaussian model predicts uniform uncertainty across all pixels. In the rightmost column, we display the signal-to-noise ratio (S/N). Interestingly, both models predict identical S/N for

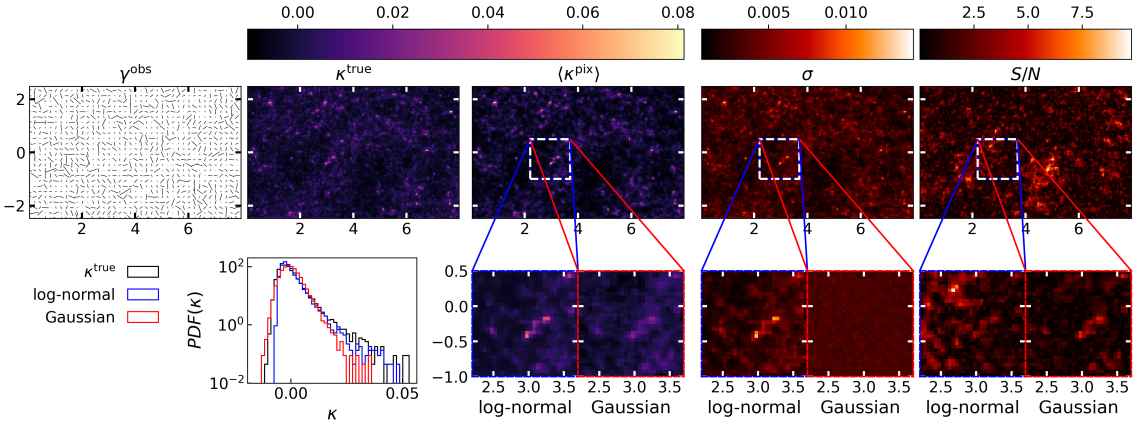

Figure 3.17: Field-level reconstruction for the first redshift bin of a single Takahashi patch using the log-normal (with informative prior on a_i) and the Gaussian models. The first row, from left to right, are the observed noisy shear map $\gamma^{\rm obs}$, the true convergence map $\kappa^{\rm true}$, the mean posterior map $\overline{\kappa^{\rm pix}}$ (log-normal model), the standard deviation map σ (log-normal model), and the S/N map (log-normal model). In the second row, the leftmost plot compares the histograms of $\kappa^{\rm true}$ with those of $\overline{\kappa^{\rm pix}}$ for both Gaussian and log-normal models. The three inset plots zoom in on $1.5\,\mathrm{deg} \times 1.5\,\mathrm{deg}$ region on the patch, contrasting the log-normal results (left) with the Gaussian results (right).

the peaks, yet the log-normal model yields a higher S/N for the voids due to its nonuniform uncertainty predictions.

For each $\kappa^{\rm pix}$ sample, we can compute its power spectrum $C(\kappa^{\rm pix})$. As proved in Zhou and Dodelson [24], in the absence of model misspecification, the distribution of $C(\kappa^{\rm pix})$ is expected to follow that of $C(\kappa^{\rm true})$. The distribution of $C(\kappa^{\rm pix})$ for the log-normal model with informative prior on a_i is depicted in Fig. 3.18, where we indeed confirm that $C(\kappa^{\rm pix})$ recovers $C(\kappa^{\rm true})$ within uncertainty.

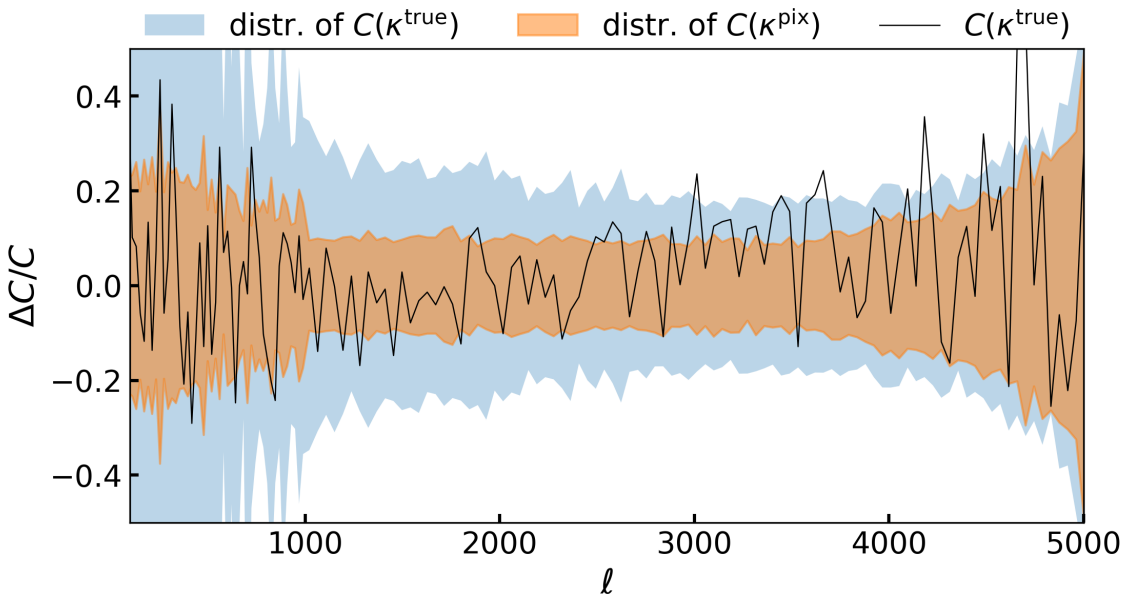

Figure 3.18: Power spectra is presented in terms of the fractional difference relative to $\overline{C(\kappa^{\text{pix}})}$, where κ^{pix} are map samples inferred from a single Takahashi patch using the log-normal model with an informative prior on a_i. The power spectrum distribution of the sampled maps, $C(\kappa^{\text{pix}})$, is depicted in orange, while the distribution of the mock fields, $C(\kappa^{\text{true}})$, is shown in blue to indicate cosmic variance. The power spectrum of the true noiseless field being inferred is shown in black. In this context, C specifies the autopower spectrum of the first redshift bin. The posterior samples are in agreement with κ^{true} at the two-point level within the bounds of uncertainty.

3.8 Summary

We have presented a percent-level accurate field-level cosmological analysis pipeline, `Miko`, that infers the joint posterior distribution of the cosmological parameters and the tomographic convergence fields. We identified several analysis systematics such as aliasing, boundary effect, mode-coupling, and density-induced shape noise, many of which have not been captured or modeled in previous studies. We measured their impact on cosmological parameter constraints and proposed effective mitigation methods that control the systematic uncertainty down to the 2% level. We then explore the problem of model misspecification (using wrong field-level priors) by testing the Gaussian and log-normal forward models on three sets of mock data (Gaussian, log-normal, and ray-tracing simulations). The Gaussian model shows no absolute bias in the A_i (the amplitude of the power spectrum in the ith redshift bin) even with model misspecification, but yields overconfident uncertainty on cosmological parameters. The log-normal provides accurate uncertainty quantification, but is sensitive to the choice of the shift parameters, a_i. In particular, the shift parameters fitted to the simulations can incur 10% absolute bias on power spectrum amplitudes. After accounting for analysis systematics (and calibrating a_i's prior for the log-normal model), both models recover A_i with an absolute bias less than 2%. Our result demonstrates the accuracy and unbiased nature of this field-level cosmological analysis framework, a prerequisite for it to provide meaningful, tight cosmological parameter constraints.

The model misspecification bias identified in this work carries important practical implications for future research. Given that the log-normal distribution results

in prior-dependent cosmological constraints, it is crucial to calibrate prior(a_i) before analyzing real data. We have demonstrated the feasibility of constructing prior(a_i) to infer unbiased cosmology from mocks. Nonetheless, there is no assurance that the same prior(a_i) will work on real data distributions. Future studies should examine the robustness of prior(a_i) across different simulations or investigate alternative one-point distributions that both reflect the non-Gaussianity of the field and remain unaffected by priors on nuisance parameters. Given the similar cosmological constraining powers derived from both the Gaussian and the log-normal models, it currently seems more practical to utilize the Gaussian model and calibrate the uncertainty estimates based on independent mock datasets.

Many analysis systematics identified in this work are applicable to general field-level inference problems involving pixelized maps. For instance, we can employ the same aliasing correction when modeling the observed temperature and polarization maps CMB at the field level. Otherwise, we risk a positive absolute bias in the inferred amplitude of the power spectrum. The error induced by model misspecification is also not weak lensing-specific.

In the next step, we will implement galaxy intrinsic alignment [106], redshift error [1], and point spread function [119] systematics into `Miko` and apply to the HSC Year 3 data. Implementing efficient algorithms to incorporate correlated pixel noise is also important [22]. Another direction is to leverage emulator-based models to constrain more general cosmological parameters [12, 120, 121].

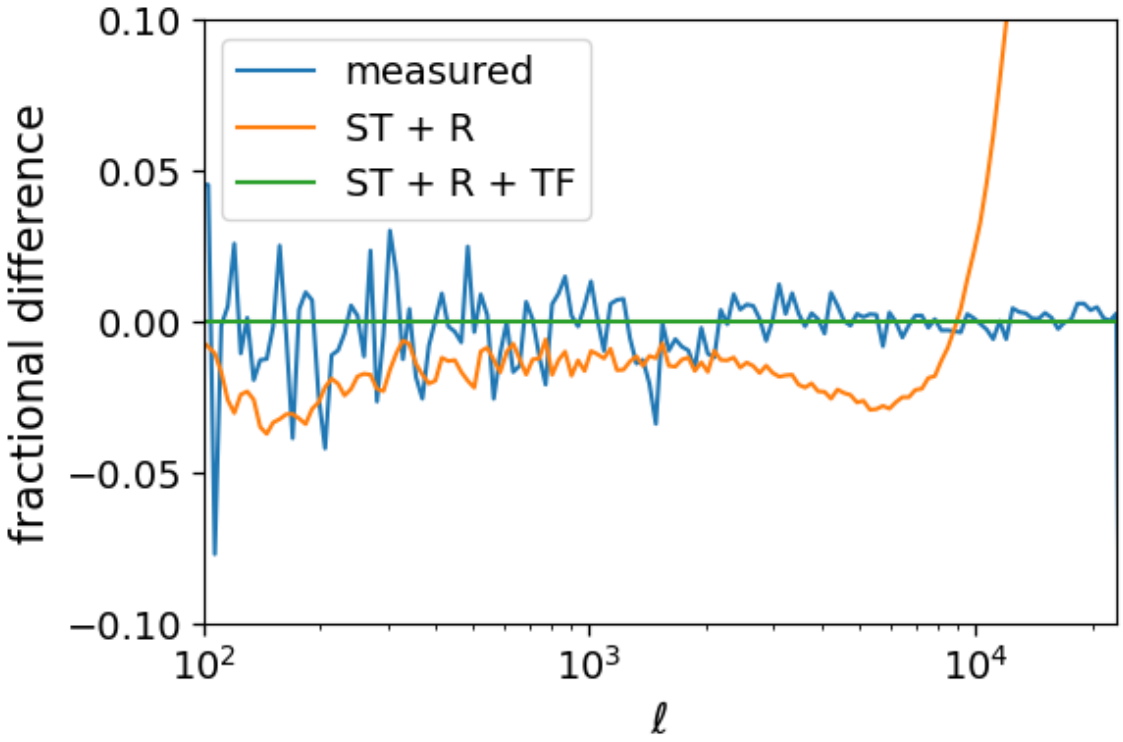

Figure 3.19: Power spectra of the first redshift bin in terms of the fractional difference relative to the adopted theoretical C for the Takahashi mocks (green). The mean of C, as measured on Takahashi mocks, is depicted in blue. The theoretical C, corrected only for shell thickness and resolution, is shown in orange. This correction is found to be insufficient for meeting our accuracy requirements. Abbreviations used: ST for shell-thickness correction, R for resolution correction, and TF for transfer function correction.

3.9 Appendix

3.10 Caliberation of the Takahashi simulations' power spectra

The convergence and shear maps calculated from the Takahashi simulations have a scale-dependent and redshift-dependent bias on the two-point level when compared to the theoretical prediction (Eq. 3.5). When we analyze the Takahashi simulations, we need to correct our theoretical model to match the simulation. At low redshift, the dominating systematics is the discretization of the lens plane. During ray tracing, the simulation assumes the light rays are deflected by lens planes of finite thickness $\Delta^{\rm ST} = 150h^{-1}$ Mpc. We therefore introduce a correction in Eq. 3.5 by convolving the $P_\delta(k;\chi)$ with a window function [108]

$$P_\delta(k;\chi) \longrightarrow \Delta^{\rm ST} \int \frac{dk_r}{2\pi} P_\delta(k;\chi) {\rm sinc}^2\left(\frac{k_r \Delta^{\rm ST}}{2}\right) \tag{3.31}$$

where k_r is the radial wave vector. At high-ℓ, the bias is mostly due to the pixelization effect. This can be corrected by [108]

$$C_{ij}(\ell) \longrightarrow \frac{C_{ij}(\ell)}{1 + (\ell/(1.6\ \texttt{nside}))^2} \tag{3.32}$$

These two corrections reduce the bias below 5% for $\ell < 10000$ as shown in orange in Fig. 3.19. Finally, we fit a transfer function so that the bias falls below 1% for $\ell < 24000$ for all redshift bins (shown in green).

CHAPTER 4

A HAMILTONIAN, POST-BORN, THREE-DIMENSIONAL, ON-THE-FLY RAY TRACING ALGORITHM FOR GRAVITATIONAL LENSING

4.1 Abstract

The analyses of the next generation cosmological surveys demand an accurate, efficient, and differentiable method for simulating the universe and its observables across cosmological volumes. We present Hamiltonian ray tracing (HRT) — the first post-Born (accounting for lens-lens coupling and without relying on the Born approximation), three-dimensional (without assuming the thin-lens approximation), and on-the-fly (applicable to any structure formation simulations) ray tracing algorithm based on the Hamiltonian formalism. HRT performs symplectic integration of the photon geodesics in a weak gravitational field, and can integrate tightly with any gravity solver, enabling co-evolution of matter particles and light rays with minimal additional computations. We implement HRT in the particle-mesh library `pmwd`, leveraging hardware accelerators such as GPUs and automatic differentiation capabilities based on `jax`. When tested on a point-mass lens, HRT achieves sub-percent accuracy in deflection angles above the resolution limit across both weak and moderately strong lensing regimes. We also test HRT in cosmological simulations on the convergence maps and their power spectra.

4.2 Introduction

Next-generation galaxy, lensing, and Cosmic Microwave Background (CMB) surveys promise percent-level or better constraints on cosmological and astrophysical parameters. This discovery potential, however, hinges on our ability to quickly and accurately simulate the universe and its observables over cosmological volumes. For example, computing covariance matrices for two-point cosmological analyses using mock catalogs requires a large number of mock catalogs [116, 122–124]. Higher-order statistical methods, such as peak and void counts, require many high-fidelity simulations to model how observables and their covariances depend on cosmology [100, 101]. Simulation-based inference (SBI) needs a large number of simulations for its training set [125]. Finally, field-level inference (FLI) demands a fast and differentiable forward model to simulate the universe and to construct observables thereof [18, 20–24, 42, 47, 87, 88, 126–134]. Perhaps not surprisingly, photons are the primary carriers of information for most cosmological observables. A photon conveys information from both its source (e.g., the temperature or polarization of the primordial CMB, or the positions of galaxies) as well as the intervening matter structures through which it traverses. In this work, we propose a new ray tracing algorithm called Hamiltonian ray tracing (HRT) that 1) accounts for the actual light trajectory and lens-lens couplings (post-Born effects in short); 2) is differentiable and computed on-the-fly; and 3) incurs minimal computational overhead when integrated with a background N-body algorithm.

The Born approximation assumes that lensing observables (e.g., weak lensing convergence κ or the CMB deflection angle) accumulate along a photon's unper-

turbed path and consider subsequent lensing events as independent from each other. Post-Born corrections effectively accounts for the ray deflection during a photons propagation and for the so-called lens-lens coupling, which describes how gravitational lenses at different redshifts can interact to generate rotational modes in the observable fields [135–137]. Post-Born effects introduce additional non-Gaussianities into lensing observables besides those generated by the non-linear clustering of the matter. Thus, it is important to include them in the modeling of higher-order statistics for both CMB and galaxy lensing. In particular, in the case of CMB lensing, non-linear and post-Born effects have similar amplitude [138, 139]. For example, post-Born effects significantly change the higher-order moments or peaks statistics of galaxy weak lensing convergence maps [140, 141]. These effects translate into cosmological biases for analysis that targets the aforementioned higher-order statistics. Moreover, post-Born corrections affect FLI through the galaxy delensing effect. As demonstrated by Ref. [142], galaxies at $z = 1$ are typically deflected on an arc minute scale, which is the typical pixel size in an FLI analysis. If the post-Born effects are unaccounted for, the FLI's forward model will generate observable fields that are coherently shifted with respect to the truth. However, since all existing lensing FLI models to date assume the Born approximation [88, 133, 143] and there does not yet exist a forward model that captures post-Born effects, the resulting cosmological biases are not well understood. Post-Born effects are less disruptive for two-point statistics, although the corrections could still be measured for the highest multipoles of the shear and lensing convergence power spectra at the level of sensitivity of future CMB lensing experiments and galaxy surveys [135, 136, 141, 144–149]. Similar con-

clusions apply for cross-correlations between CMB lensing and galaxy survey probes [150].

Weak lensing simulations typically capture post-Born effects using the multiple lens plane (MLP) formalism [145, 151–153], which has been widely applied in many studies [108, 139, 149, 154–162] Our new HRT formalism improves upon the MLP method in ways that enable new cosmological analysis methods. **First**, whereas MLP deflects photons only by the gravitational potential gradients instantaneously at a lens plane, HRT utilizes the gravitational potential of the entire simulated volume without the thin-lens approximation, a feature known as three-dimensional ray tracing [151, 163–166]. **Second**, the MLP algorithm requires initially running an N-body simulation, storing its time steps (called snapshots), preparing the density field on the light cone in mass shells, and finally accumulating the convergence fields backward in time. This multi-step process decouples ray tracing from the N-body simulations, complicating implementation when varying cosmology and introducing computational overhead in terms of data storage and transfer. Our HRT method performs ray tracing on-the-fly by co-evolving the light rays with the matter particles. The physics and the time-stepping of the rays are tightly synchronized with the N-body simulation and subject to the same cosmological model, thereby eliminating the need for data storage. Refs. [164, 165]'s algorithm performs three-dimensional ray tracing on-the-fly, but ignore post-Born effects and lack differentiability and other features listed below. **Third**, FLI requires that the observable fields be differentiable with respect to all cosmological parameters and the initial conditions of the background N-body simulation. Partly because of the multi-step process, current MLP

codes are not differentiable. HRT formalism leverages a combination of automatic differentiation routines in `jax` and variational equation to achieve efficient differentiation with respect to all parameters of interest. **Fourth**, MLP typically uses Kaiser-Squires algorithm [145] or finite-difference [153] methods to compute lensing observables, which could suffer from numerical instability. As we shall see, since most lensing observables are defined in terms of derivatives of the photon trajectory with respect to its initial position, and our new methods are differentiable, we can calculate these observables by direct differentiation. **Fifth**, unlike MLP which approximates the total lensing deflection as a Riemann sum over the radial comoving space, HRT solves each photon's (light ray's) equations of motion (EOMs) using a highly stable and accurate symplectic integrator. This provides a systematic way to correct for the finite time-resolution of the snapshots. **Finally**, MLP implementations do not currently leverage hardware accelerators while HRT is implemented on both CPUs and GPUs. Moreover, on-the-fly ray-tracing is particularly suited for GPU applications since GPU-based N-body simulations are already memory limited and cannot afford to store snapshots and post-process them later into shear maps.

The paper is organized in the following way. We first review the Hamiltonian formulation of the photon dynamics in a gravitational field in Section 4.3. Next, we discretize the EOMs to construct a symplectic integrator for the photons' trajectories in Section 4.4. Once we have the trajectories, we construct an algorithm that calculates the weak lensing observables along the light path in Section 4.5. We discuss an efficient implementation of the algorithm in Section 4.6, and test the algorithm for point mass deflection and weak lensing maps in cosmological volumes in Section 4.7.

We conclude in Section 4.8.

4.3 Hamiltonian dynamics of light rays

In post-Born ray tracing, the trajectories of photons are deflected by the large-scale structure. Consequently, it is not known a priori if a photon emitted by a source will reach the observer. The core idea of HRT is to evolve both light rays (from its observed position on the image plane, denoted by $\boldsymbol{\theta}_0$) and the matter particles (from their displacements and velocities at today) backward in time in an N-body simulation[1] The light ray EOMs are reversible once we describe them with Hamiltonian dynamics, which we then implement numerically using symplectic integration. This approach involves three parts: constructing the spacetime metric with gravitational perturbations, defining the photon's Hamiltonian using this metric, and finally, solving Hamilton's equations. We only motivate the results in this section and leave the detailed derivations to Section 4.9.1.

Let us parameterize spacetime with conformal time τ and comoving spatial coordinates $\mathbf{x}$. On scales smaller than the cosmic horizon, the matter density contrast $\delta(\mathbf{x}, \tau)$ sources a gravitational potential field $\Phi(\mathbf{x}, \tau)$ via Poisson's equation,

$$\boldsymbol{\nabla}^2 \Phi(\mathbf{x}, \tau) = \frac{3}{2} \frac{\Omega_{\mathrm{m}} H_0^2}{a(\tau)} \delta(\mathbf{x}, \tau), , \tag{4.1}$$

1. We can backward evolve the dark matter particles by reversing the time variable in their EOMs. Table 2 of Ref. [126] has shown that both the positions and velocities of the dark matter particles can be recovered to high accuracy across different simulation configurations in `pmwd` when the EOMs are solved with a symplectic integrator.

where the gradient (unless specified) is with respect to the comoving coordinates. Here, a, $\Omega_{\rm m}$, and H_0 represent the scale factor, matter fraction, and Hubble parameter, respectively. The potential $\Phi(\mathbf{x}, \tau)$ perturbs the otherwise homogeneous and isotropic Friedmann-Lemaître-Robertson-Walker metric, given by

$$ds^2 = -a^2\Big(1 + 2\frac{\Phi}{c^2}\Big)d\tau^2 + a^2 h_{ij} dx^i dx^j \tag{4.2}$$

$$\equiv -a^2\Big(1 + 2\frac{\Phi}{c^2}\Big)d\tau^2 + a^2\Big(1 - 2\frac{\Phi}{c^2}\Big)(r^2 d\boldsymbol{\theta}^2 + d\chi^2),, \tag{4.3}$$

for $i = 1, 2, 3$. r and χ are the transverse and radial comoving distances, respectively, and $r = \chi$ in a flat universe. We parameterize the sky under the flat-sky approximation with angular coordinates, i.e., $\mathbf{x} = (\boldsymbol{\theta}, \chi)$.

Now, consider a photon. The general relativistic covariant Hamiltonian of a photon is given by

$$H(x^j, p_i, \tau) = c\sqrt{h^{ij} p_i p_j}\Big(1 + 2\frac{\Phi}{c^2}\Big),, \tag{4.4}$$

where p_i is the momentum one-form [167]. The EOMs of the photon are solutions to Hamilton's equations. The results are (see detailed proof and discussion in Section 4.9.1; Ref. [167] also derives a similar result albeit with a different parameterization of the image plane)

$$\frac{d\boldsymbol{\theta}}{d\chi} = -\frac{\boldsymbol{\eta}}{r^2}, \tag{4.5}$$

$$\frac{d\boldsymbol{\eta}}{d\chi} = \frac{2}{c^2}\frac{\partial\Phi}{\partial\boldsymbol{\theta}}, \tag{4.6}$$

$$\frac{d\tau}{d\chi} = -\frac{1}{c}\left(1 - 2\frac{\Phi}{c^2} + \frac{\boldsymbol{\eta}^2}{2r^2}\right), \tag{4.7}$$

where we have defined the conjugate momentum

$$\boldsymbol{\eta} = \frac{r\boldsymbol{v}}{c},, \tag{4.8}$$

and $\boldsymbol{v}$ is the *transverse* peculiar velocity of the photon. Equation (4.5) is simply a geometric relation in the deflection tangent: $-rd\boldsymbol{\theta}/d\chi = \boldsymbol{v}/c$. The second and third terms in Eq. (4.7) correspond to the Shapiro and the geometric time delay, respectively. $\boldsymbol{\theta}$ has the unit of angle while $\boldsymbol{\eta}$ has the unit of length[2]. The initial conditions are given by

$$\boldsymbol{\theta}(\chi_0) = \boldsymbol{\theta}_0,, \quad \boldsymbol{\eta}(\chi_0) = 0,, \quad \chi_0 = 0,. \tag{4.9}$$

To arrive at these EOMs, we have used the flat-sky (small angle) approximation, kept Φ/c^2, $\boldsymbol{\nabla}\Phi/c^2$, and assumed $(\boldsymbol{v}/c)^2$ is of $\mathcal{O}(\Phi/c^2)$. We have also used the comoving distances instead of conformal time to parameterize the photon trajectory. The two are related by Eq. (4.7), where the second term is the Shapiro time delay. However, we will not consider the time delay correction in our study.

2. They are actually conjugate variables in a separable Hamiltonian and relates to the Fermat action as shown in Section 4.9.2. In fact, the time delay terms can also be seen easily from the separable Hamiltonian.

4.4 Symplectic ray tracing

4.4.1 Equations of motion in discrete time

So far, we have derived the EOMs for photons in angular coordinates and conjugate momenta using the Hamiltonian formalism. We now introduce a second-order symplectic integrator to integrate these EOMs and trace the photons' trajectories. A symplectic integrator advances the equations of motion in discrete time steps, each of which preserves the mathematical structure of the Hamiltonian. As such, a symplectic integrator maintains the long-term behavior of the particle's motion. Symplectic integrators have been widely used in cosmological simulations and achieve superior accuracy compared to other integration methods of the same or even higher orders [168–170].

In this work, we utilize the kick-drift-kick (KDK) integrator, also known as the velocity Verlet method [171]. We will only sketch out the key ideas here, and leave a more detailed analysis of the integrator to Section 4.9.2. The KDK integration scheme operates by decomposing each integration step of the EOMs, $\Delta\chi$, into three consecutive stages, updating either the positions or the momenta at a time. Initially, we update $\boldsymbol{\eta}$ for a half time step ($\Delta\chi/2$), leaving $\boldsymbol{\theta}$ unchanged (the first kick). Next, we update $\boldsymbol{\theta}$ for a full time step ($\Delta\chi$), while keeping $\boldsymbol{\eta}$ unchanged (the drift stage). Finally, we apply another half time step update to $\boldsymbol{\eta}$ to complete the integration (the second kick).

To implement this integration scheme, we configure the time steps in our simulations as follows. We perform ray tracing while evolving the N-body simulation

backward in time. Light is traced from the observer's position at $\chi = 0$ and $\tau = \tau_0$, towards the light source at $\chi_s > 0$ and $\tau = \tau_0 - \chi_s/c$. We define a series of lens in mass shells perpendicular to the main line of sight, labeled by the subscript n, with $n = 0$ at the observer and $n = s$ at the source. The lens shells are determined by the time stepping of the forward simulation itself, with each shell sandwiched between the light fronts at two consecutive time steps, and the n-th shell's midpoint labeled by $n + 1/2$. With this set up, we can now integrate the EOMs iteratively via (see Section 4.9.2 for derivation and error analysis)

$$\boldsymbol{\eta}_{n+\frac{1}{2}} = \boldsymbol{\eta}_n + K_n^{n+1/2}(\boldsymbol{\theta}_n, \chi_n),, \tag{4.10}$$

$$\boldsymbol{\theta}_{n+1} = \boldsymbol{\theta}_n + D_n^{n+1}(\boldsymbol{\eta}_{n+1/2}, \chi_{n+1/2}),, \tag{4.11}$$

$$\boldsymbol{\eta}_{n+1} = \boldsymbol{\eta}_{n+1/2} + K_{n+1/2}^{n+1}(\boldsymbol{\theta}_{n+1}, \chi_{n+1}),, \tag{4.12}$$

where

$$K_a^b(\boldsymbol{\theta}_c, \chi_c) = -\frac{2}{c}\int_{\chi_a}^{\chi_b} d\chi, r(\chi)\boldsymbol{\nabla}_\perp \Phi(\mathbf{x}_c, \tau(\chi)),, \tag{4.13}$$

$$D_a^b(\boldsymbol{\eta}_c, \chi_c) \simeq (\frac{\boldsymbol{\eta}_c}{cr_c^2})(\chi_b - \chi_a),, \tag{4.14}$$

are the kick and the drift factors, and $\boldsymbol{\nabla}_\perp \equiv r^{-1}\boldsymbol{\nabla}_{\boldsymbol{\theta}}$ denotes the gradient with respect to the transverse comoving coordinates. Note that the traditional velocity Verlet algorithm typically approximates the first kick operator in Equation (4.10) as $\frac{1}{2}K_{n-1/2}^{n+1/2}(\boldsymbol{\theta}_n, \chi_n)$. The proposed iteration thus provides a better comoving space (time) resolution.

4.4.2 Time resolution correction

Equation (4.13) updates light ray's conjugate momenta by integrating the *time-evolving* $\boldsymbol{\nabla}_\perp \Phi(\mathbf{x}, \tau(\chi))$ along their trajectories. In practice, we derive $\boldsymbol{\nabla}_\perp \Phi(\mathbf{x}, \tau(\chi))$ from the forces in N-body simulation, at discrete time steps often referred to as "snapshots", say at $\tau(\chi_c)$. If each time step is small, we can approximate the growth of the matter structure using

$$\Phi(\mathbf{x}, \tau(\chi)) = \Phi(\mathbf{x}, \tau_c) \frac{D_1(\tau)}{D_1(\tau_c)} \frac{a(\tau_c)}{a(\tau)}, , \tag{4.15}$$

where D_1 is the linear growth function, $\tau = \tau(\chi)$, and $\tau_c = \tau(\chi_c)$. Using this correction, Eq. (4.13) becomes

$$K_a^b(\boldsymbol{\theta}_c, \chi_c) = -\frac{2}{c} \int_{\chi_a}^{\chi_b} d\chi, r(\chi) \boldsymbol{\nabla}_\perp \Phi(\mathbf{x}_c, \tau) \frac{D_1(\tau)}{D_1(\tau_c)} \frac{a(\tau_c)}{a(\tau)}, . \tag{4.16}$$

4.5 Weak lensing observables

In Section 4.4, we worked out the trajectories of the photons. We now aim to construct observable maps on the image plane using these photon trajectories. Most cosmological observables O can be written as line-of-sight integrals along the light path,

$$O(\boldsymbol{\theta}) = \int_{\boldsymbol{\theta}(\chi)} d\chi, P(\chi) Q(\mathbf{x}, \tau(\chi)), , \tag{4.17}$$

where P is the line-of-sight kernel, and Q is any cosmological field. Examples include weak lensing distortions, the galaxy density, the Sunyaev-Zel' dovich effects, the

integrated Sachs-Wolfe effect, and the dispersion measure [5, 165, 172]. Here, we focus on weak lensing and develop an efficient and accurate algorithm for computing the convergence, shear, and rotation maps.

We begin with the distortion matrix $\mathbf{A}$, which characterizes the lensing effect of an object at source χ,

$$\mathbf{A}(\chi) = \frac{\partial \boldsymbol{\theta}(\chi)}{\partial \boldsymbol{\theta}_0},. \tag{4.18}$$

$\mathbf{A}$ can be decomposed into the product of a rotation and a shear matrix[3]:

$$\mathbf{A} = \begin{pmatrix} \cos\omega & \sin\omega \\ -\sin\omega & \cos\omega \end{pmatrix} \begin{pmatrix} 1-\kappa-\gamma_1 & -\gamma_2 \\ -\gamma_2 & 1-\kappa+\gamma_1 \end{pmatrix},. \tag{4.19}$$

The convergence κ, shear γ, and rotation ω are the primary weak lensing observables. Since ω is orders of magnitude smaller than κ and γ, we solve for the observables in terms of the entries of $\mathbf{A}$ keeping ω to first order,

$$\begin{cases} \kappa = 1 - C_1,, & \gamma_1 = C_3 + C_4,\omega,, \\ \omega = \dfrac{C_2}{1-\kappa},, & \gamma_2 = -C_4 + C_3,\omega,, \end{cases} \tag{4.20}$$

where

$$\begin{cases} C_1 = \dfrac{A_{11}+A_{22}}{2},, & C_3 = \dfrac{A_{22}-A_{11}}{2},, \\ C_2 = \dfrac{A_{12}-A_{21}}{2},, & C_4 = \dfrac{A_{12}+A_{21}}{2},. \end{cases} \tag{4.21}$$

Finally, going back to Eq. (4.17), the line-of-sight kernel P is the normalized galaxy

3. omitting χ for clarity

density distribution $n_{\rm g}(\chi)$, and Q is either κ, γ, or ω.

Past works have computed these observables using either the Kaiser-Squires inversion [145] or finite differences [153, 165]. However, the former is sensitive to boundary conditions and the latter can induce numerical instability. We propose a new algorithm that computes the observable maps via direct (forward-mode) differentiation of Eq. (4.18), which is convenient to implement in our framework using `jax` automatic differention.

The main idea is to decompose A_n[4] as chained products of Jacobian matrices for time steps $n' < n$. Differentiating the discretized EOM in Eq. (4.11) with respect to the initial ray position using the chain rule,

$$\begin{aligned}\frac{\partial \boldsymbol{\theta}_{n+1}}{\partial \boldsymbol{\theta}_0} &= \frac{\partial \boldsymbol{\theta}_n}{\partial \boldsymbol{\theta}_0} + \frac{\partial D_n^{n+1}}{\partial \boldsymbol{\eta}_{n+1/2}} \frac{\partial \boldsymbol{\eta}_{n+1/2}}{\partial \boldsymbol{\theta}_0} \\ &= \frac{\partial \boldsymbol{\theta}_n}{\partial \boldsymbol{\theta}_0} + \frac{\partial D_n^{n+1}}{\partial \boldsymbol{\eta}_{n+1/2}} \left(\frac{\partial \boldsymbol{\eta}_n}{\partial \boldsymbol{\theta}_0} + \frac{\partial K_n^{n+1/2}}{\partial \boldsymbol{\theta}_n} \frac{\partial \boldsymbol{\theta}_n}{\partial \boldsymbol{\theta}_0} \right) \end{aligned} \tag{4.22}$$

Using the helper function

$$\boldsymbol{B}(\chi) = \frac{\partial \boldsymbol{\eta}(\chi)}{\partial \boldsymbol{\theta}_0}, , \tag{4.23}$$

4. We use the subscript n to denote observables computed at χ_n, e.g., $A(\chi_n) = A_n$.

we can then rewrite Eq. (4.22) as an iteration,

$$\boldsymbol{B}_{n+1/2} = \boldsymbol{B}_n + \frac{\partial K_n^{n+1/2}}{\partial \boldsymbol{\theta}_n} \mathbf{A}_n, , \tag{4.24}$$

$$\mathbf{A}_{n+1} = \mathbf{A}_n + \frac{\partial D_n^{n+1}}{\partial \boldsymbol{\eta}_{n+1/2}} \boldsymbol{B}_{n+1/2}, , \tag{4.25}$$

$$\boldsymbol{B}_{n+1} = \boldsymbol{B}_{n+1/2} + \frac{\partial K_{n+1/2}^{n+1}}{\partial \boldsymbol{\theta}_{n+1}} \mathbf{A}_{n+1}, , \tag{4.26}$$

with the initial conditions (for each photon),

$$\mathbf{A}_0 = \mathbb{I}_{2\times 2}, , \quad \boldsymbol{B}_0 = \mathbb{O}_{2\times 2}, . \tag{4.27}$$

Here, $\mathbb{I}$ and $\mathbb{O}$ are the identity and the zero matrices, respectively. We compute the iterations for $\mathbf{A}$ and $\boldsymbol{B}$ while calculating the trajectories of the photons. Each $\mathbf{A}$ and $\boldsymbol{B}$ iteration step is a Jacobian-vector product, allowing us to calculate them automatically using the forward differentiation in `jax`, with minimal computational and memory overhead. Finally, we accumulate the observables by

$$O = \sum_n n_{\mathrm{g}}(\chi_n), O_n, , \tag{4.28}$$

where $O_n \in \{\kappa_n, \gamma_n, \omega_n\}$.

Algorithm 1 Reverse-time co-evolution of dark matter and a single ray

procedure RAYTRACING($\mathcal{P}_\mathrm{P}, , \mathcal{P}_\mathrm{DM}, , n_g$)
  $A \leftarrow \mathbb{I}_{2\times 2}, , B \leftarrow \mathbb{O}_{2\times 2}$
  $[\kappa, , \gamma_1, , \gamma_2, , \omega] \leftarrow \mathbf{0}$
  $a \leftarrow 1$
  while $a \geq a_\mathrm{min}$ **do** ▷ ray-trace backward in time to a_min
    $\boldsymbol{\nabla}_\perp \Phi \leftarrow \mathcal{P}_\mathrm{DM}$
    $\mathcal{P}_\mathrm{DM} \leftarrow \texttt{nbody_reverse_step}(\mathcal{P}_\mathrm{DM}, , \boldsymbol{\nabla}_\perp \Phi)$ ▷ Evolve dark matter backward [126]
    $\mathcal{P}_\mathrm{P} \leftarrow \texttt{kick}(\mathcal{P}_\mathrm{P}, , \boldsymbol{\nabla}_\perp \Phi)$ ▷ Eq. (4.5)
    $B \leftarrow \texttt{iterate_B}(\mathcal{P}_\mathrm{P}, , A, , B, , \boldsymbol{\nabla}_\perp \Phi)$ ▷ Eq. (4.24)
    $\mathcal{P}_\mathrm{P} \leftarrow \texttt{drift}(\mathcal{P}_\mathrm{P})$ ▷ Eq. (4.6)
    $A \leftarrow \texttt{iterate_A}(\mathcal{P}_\mathrm{P}, , A, , B)$ ▷ Eq. (4.25)
    $\mathcal{P}_\mathrm{P} \leftarrow \texttt{kick}(\mathcal{P}_\mathrm{P}, , \boldsymbol{\nabla}_\perp \Phi)$ ▷ Eq. (4.5)
    $B \leftarrow \texttt{iterate_B}(\mathcal{P}_\mathrm{P}, , A, , B, , \boldsymbol{\nabla}_\perp \Phi)$ ▷ Eq. (4.26)
    $\kappa, , \gamma_1, , \gamma_2, , \omega \leftarrow \texttt{observe}(A, , n_g, , \kappa, , \gamma_1, , \gamma_2, , \omega)$ ▷ Eqs. (4.20) and (4.28)
    $a \leftarrow a - \Delta a$
  end while
  return $\kappa, , \gamma_1, , \gamma_2, , \omega$
end procedure

4.6 Implementation

4.6.1 Efficient dark matter and photons co-evolution

In this section, we propose an implementation of the HRT algorithm that co-evolves the dark matter particles and light rays[5]. This implementation only incurs minimal computational overhead compared to evolving the dark matter particles alone. The HRT algorithm is applicable to all numerical simulations of structure formation, and particularly so to differentiable ones. Specifically, we have implemented it atop the

5. Each ray (bundle) is defined as the collection of photons observed in the same pixel on the image plane.

`pmwd` library, which offers a differentiable, fast, and memory-efficient particle mesh-based dark matter simulation [126, 127].

Let us first evolve a dark matter distribution from the initial condition to $a = 1$ where the observer is. To start ray tracing, we first initialize the rays on a uniform grid representing the pixels (of size μ_{2D}; Table 4.1 lists the mesh variables in this section) on the image plane at $a = 1$. To simplify our discussion, we will focus on tracing a single ray. The actual implementation parallelizes trivially across all the rays since there are no interactions between them. We use $\mathcal{P}_{\mathrm{P}}$ to denote the state (position and momentum) of this ray and $\mathcal{P}_{\mathrm{DM}}$ to represent the state of all the dark matter particles. We co-evolve the matter particles and the light rays backward in time as described in Algorithm 1. At each time step, we compute the potential gradient $\boldsymbol{\nabla}_{\perp}\Phi$ on the 3D mesh. We then use $\boldsymbol{\nabla}_{\perp}\Phi$ to first evolve the dark matter particles (`nbody_reverse_step` in Algorithm 1, [126]) backward in time and then to integrate the ray's EOMs via the KDK integrator (Section 4.4).

Table 4.1: Definitions of mesh and ray variables. The white rows define the resolution of meshes or rays, while the light gray ones define their corresponding position vectors. The top two rows are for the 3D particle mesh, on which the gravitational forces are computed. The middle two rows specify the 2D configuration of rays. And the bottom two rows describe the 2D ray mesh on which we interpolate and transfer the transverse gravitational forces, from the 3D particle mesh and to the 2D ray positions.

Variable	**Purpose / Definition**
l_{3D}	3D particle mesh comoving resolution
$\mathbf{x}_{\mathrm{3D}}$	3D positions of particle mesh grid points
μ_{2D}	2D ray spacing / pixel resolution on the image plane
$\boldsymbol{\theta}_0, \boldsymbol{\theta}$	2D ray positions on the image plane and during ray tracing, respectively
ν_{2D}	2D ray mesh resolution during ray tracing
$\boldsymbol{\theta}_{\mathrm{2D}}$	2D positions of ray mesh grid points during ray tracing

The most computationally expensive operation in the simulation is to calculate $\boldsymbol{\nabla}_\perp \Phi$ via the Fast Fourier Transform (FFT) [126]. Algorithm 1, however, only requires $\boldsymbol{\nabla}_\perp \Phi$ to be calculated once per time step. The extra computation is either in 2D or on a thin shell of 3D field (e.g., the computation of the kick operator as detailed below in Section 4.6.2), and thus is negligible compared to the 3D work load already done by the gravity solvers.

4.6.2 The kick operator

In a particle mesh (PM)-based N-body simulation, we first compute the gravitational force field on a 3D mesh and then interpolate this force onto particle positions. Similarly, we compute the force on light rays (in the kick operator, $K_a^b(\boldsymbol{\theta}_c, \chi_c)$ in Eq. (4.16), where $\boldsymbol{\theta}_c$ is the position of the ray at the integration step) using the PM method. Our algorithm achieves this using 3 meshes/arrays: a 3D particle mesh covering the entire simulation volume with resolution $l_{\rm 3D}$ and coordinates $\mathbf{x}_{\rm 3D}$, a 2D array of rays covering the image plane with angular resolution $\mu_{\rm 2D}$ and coordinates $\boldsymbol{\theta}_0$, and an intermediate 2D ray mesh transferring forces from the first mesh to the second with angular resolution $\nu_{\rm 2D}$ and coordinates $\boldsymbol{\theta}_{\rm 2D}$.

We first evaluate the integrand of Eq. (4.16) on the 3D mesh. For mesh points within the lens plane (having comoving coordinates $\chi_a \leq \chi < \chi_b$), we calculate their projected angles, then interpolate and accumulate their values onto the ray mesh using cloud-in-cell (CIC), or trilinear, interpolation [173]. This process evaluates the kick operator line-of-sight integration on the ray mesh, $K_a^b(\boldsymbol{\theta}_{\rm 2D}, \chi_c)$. To account for the smoothing effect introduced by the CIC interpolation, we deconvolve $K_a^b(\boldsymbol{\theta}_{\rm 2D}, \chi_c)$

in Fourier space as follows,

$$\tilde{K}_a^b(\mathbf{k}, \chi_c) \to \tilde{K}_a^b(\mathbf{k}, \chi_c)\ \mathrm{sinc}\Big(\frac{\mathbf{k}_x \nu_{\mathrm{2D}}}{2}\Big)^{-4}\ \mathrm{sinc}\Big(\frac{\mathbf{k}_y \nu_{\mathrm{2D}}}{2}\Big)^{-4},, \tag{4.29}$$

where $\tilde{\cdot}$ denotes the Fourier space quantity and $\mathbf{k}$ is the wave vector. And the powers of -4 on the sinc functions accounts for both the interpolation from the 3D particle mesh to the 2D ray mesh, and that from the latter to the rays. To account for the finite width of the ray bundle, we apply Gaussian smoothing on $K_a^b(\boldsymbol{\theta}_{\mathrm{2D}}, \chi_c)$ at the resolution

$$\lambda_{\mathrm{lim}} = \max\Big(\frac{2, l_{\mathrm{3D}}}{r_a + r_b}, \mu_{\mathrm{2D}}\Big),, \tag{4.30}$$

by applying

$$\tilde{K}_a^b(\mathbf{k}, \chi_c) \to \tilde{K}_a^b(\mathbf{k}, \chi_c) \exp\Big(\frac{1}{2}(\mathbf{k}_x^2 + \mathbf{k}_y^2)\lambda_{\mathrm{lim}}^2\Big),. \tag{4.31}$$

Finally, we use CIC interpolation to map $K_a^b(\boldsymbol{\theta}_{\mathrm{2D}}, \chi_c)$ onto the positions of the rays, resulting in $K_a^b(\boldsymbol{\theta}_c, \chi_c)$.

4.6.3 Adaptive ray mesh

Unlike the 3D mesh which has a fixed comoving resolution, an angle on the 2D mesh corresponds to a comoving length that varies with the redshift. For example, at low z, $\nu_{\mathrm{2D}}\chi(z) \ll l_{\mathrm{3D}}$. As a result, a 2D ray mesh samples forces from the 3D particle mesh at a much higher resolution the latter doesn't offer. So maintaining a high resolution 2D mesh at low redshift is both computationally inefficiency. Therefore, we employ an adaptive ray mesh, where we progressively coarsen ν_{2D} towards low

redshifts. The choice of ν_{2D} depends on the 3D mesh resolution, pixel size, and efficient FFT sizes (including padding). Further details and convergence tests are discussed in Section 4.9.3.

4.7 Validations

4.7.1 Lensing by a point mass

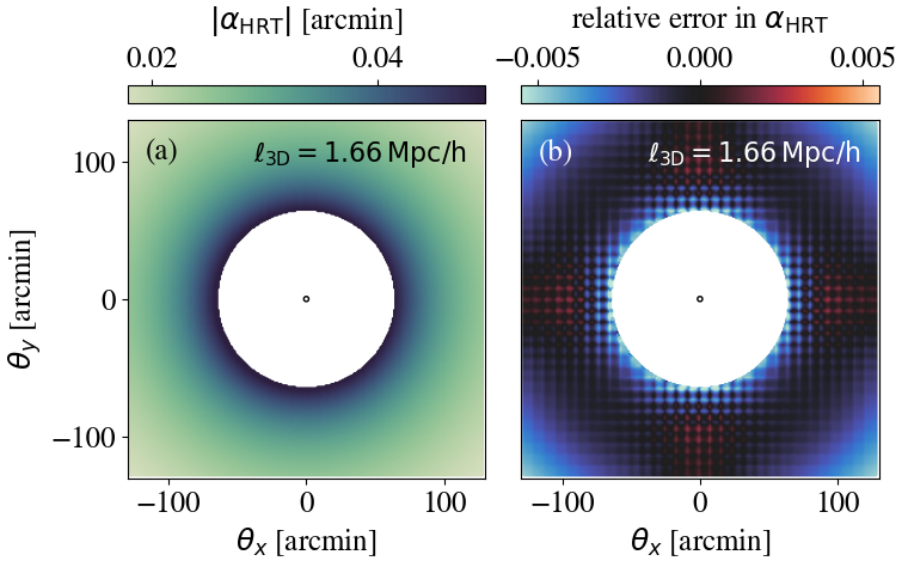

Figure 4.1: We apply HRT to the problem of point mass lensing as described in Section 4.7.1, using a $256 \times 256 \times 512$ box with a mesh resolution $l_{3D} = 1.66, \mathrm{Mpc}/h$. We show $\boldsymbol{\alpha}_{\mathrm{HRT}}$ in panel (a). For a given l_{3D}, the point mass angular resolution limit is $\theta_{\mathrm{lim}} = 4, l_{3D}/\chi_{\mathrm{l}}$ (see Fig. 4.2). We mask light rays within this threshold where the potential field of the point mass lens is not well resolved. The black circle indicates the Einstein radius θ_{E}. Panel (b) illustrates the relative error between $\boldsymbol{\alpha}_{\mathrm{HRT}}$ and $\boldsymbol{\alpha}_{\mathrm{theory}}$. HRT achieves accuracy within $< 0.5\%$ for both configurations above the resolution limit.

We test our ray tracing algorithm on the classic problem of gravitational lensing

by a point mass in a flat, static universe. The observer is positioned at $\chi = 0$, with a lens mass $M = 1.3 \times 10^{15} M_\odot$ (with an Einstein radius of $\theta_{\rm E} = 1.8'$, representative of a massive cluster) at $\chi_{\rm l} = 350$ Mpc$/h$, and the source plane at $\chi_{\rm s} = 850$ Mpc$/h$ (or $z_{\rm s} \approx 0.3$). The theoretical prediction for the deflection angle $\boldsymbol{\alpha}_{\rm theory} \equiv \boldsymbol{\theta}_{\rm theory} - \boldsymbol{\theta}_0$ is provided by [174]:

$$\boldsymbol{\alpha}_{\rm theory} = \frac{\chi_{\rm s} - \chi_{\rm l}}{\chi_{\rm s}} \Big[\hat{\alpha} + \frac{15\pi}{4} \hat{\alpha}^2 + \mathcal{O}\left(\hat{\alpha}^3\right) \Big] \frac{\boldsymbol{\theta}_0}{|\boldsymbol{\theta}_0|}, , \tag{4.32}$$

where $\hat{\alpha} \equiv 4GM/(bc^2)$ and $b = \chi_{\rm l} |\boldsymbol{\theta}_0|$ is the impact parameter. We solve the same problem using HRT. We define a 3D mesh of size $(256, 256, 512)$ with a resolution of 1.66 Mpc$/h$, and an image plane spanning $256' \times 256'$ with a pixel resolution of $1'$. We initialize rays on a uniform grid ($\boldsymbol{\theta}_0$) at $z = 0$ and trace them to $z_{\rm s}$ in 45 time steps. Panel (a) of Fig. 4.1 shows the deflection angle $\boldsymbol{\alpha}_{\rm HRT}$ obtained via HRT.

The HRT result aligns with theoretical predictions with very high accuracy. We plot the relative error on the image plane, defined by $\epsilon = |\boldsymbol{\alpha}_{\rm HRT} - \boldsymbol{\alpha}_{\rm theory}|/|\boldsymbol{\alpha}_{\rm theory}|$, in panel (b) of Fig. 4.1. HRT consistently achieves accuracy within 0.5% across the image plane where the 3D mesh resolution is adequate. Specifically, we mask pixels within $\theta_{\rm lim} = 4, l_{\rm 3D}/\chi_{\rm l}$, where the potential field generated by the point particle cannot be clearly resolved due to the finite resolution of the 3D mesh. A lower resolution mesh dampens the potential field at the mesh scale, thereby suppressing $\boldsymbol{\alpha}_{\rm HRT}$ near the lens mass. Figure 4.2 illustrates ϵ for the $M = 1.3 \times 10^{15} M_\odot$ case above as a function of $\theta_{\rm lim}$ for two configurations with different resolutions: $l_{\rm 3D} =$ 1.66 Mpc$/h$ and 3.32 Mpc$/h$. These two configurations test HRT in the weak field limit, where the Einstein radius $\theta_{\rm E} \ll \theta_{\rm lim}$. As expected, for pixels falling within

$\theta_{\rm lim}$, $\boldsymbol{\alpha}_{\rm HRT}$ is systematically lowered (Fig. 4.2, solid blue line). Halving the 3D mesh resolution doubles $\theta_{\rm lim}$, but $\boldsymbol{\alpha}_{\rm HRT}$ still demonstrates percent-level accuracy outside $\theta_{\rm lim}$ and is suppressed within it (Fig. 4.2, solid orange line). We also test HRT in a stronger gravitational field by increasing M such that $\theta_{\rm E} \approx \theta_{\rm lim}$. The result for the high and low resolution cases are shown in dashed green and red lines in Fig. 4.2, respectively. In general, we find the definition of $\theta_{\rm lim}$ serves as a robust and universal threshold to characterize the accuracy of HRT regardless of $l_{\rm 3D}$ and M. This also shows that HRT is accurate as long as the 3D mesh resolution is sufficiently high. The accuracy of HRT also weakly depends on the resolution and boundary conditions of the ray mesh, as well as the number of time steps. We characterize these effects in Section 4.9.4.

4.7.2 Post-Born weak lensing in a cosmological volume

We use the HRT algorithm to perform ray tracing over a cosmological volume and study the statistics of weak lensing observables. Throughout this paper, we adopt the *Planck* 2015 cosmology [175]. This cosmology also underpins the κTNG simulations [157] and the lensing forward model by Ref. [133], with which we compare our results. The κTNG simulations are obtained by post-processing the higher resolution dark matter-only TNG300-1-Dark simulations (hereafter TNG-Dark) using the MLP ray tracing algorithm [157, 176–180]. Ref. [133] constructed weak lensing convergence maps with PM simulation using the Born approximation. The main purpose of this paper is to present the HRT algorithm itself; we will focus on power spectrum recovery here and leave detailed higher-order statistics analysis for a future work.

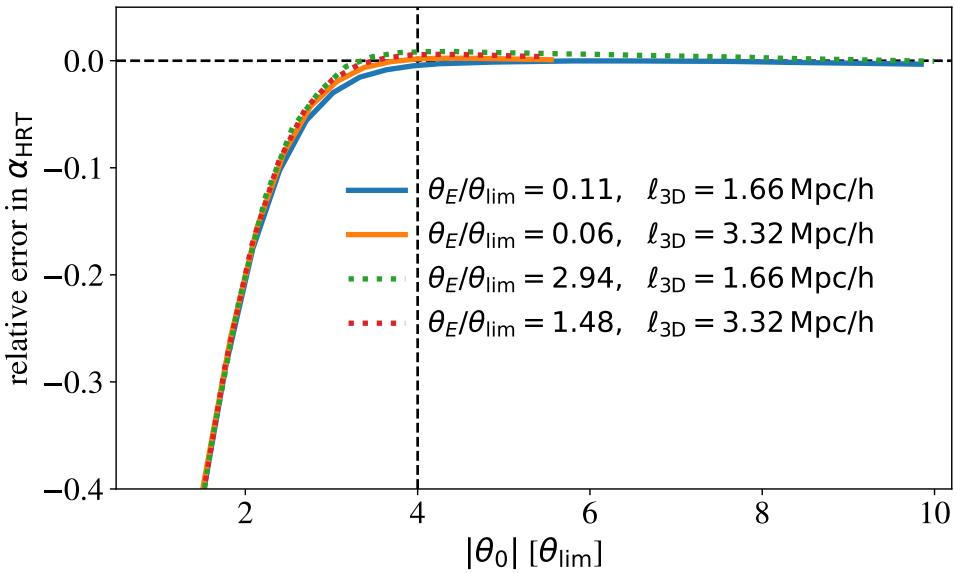

Figure 4.2: Relative error of $\boldsymbol{\alpha}_{\rm HRT}$ versus $|\theta_0|$ (in unit of $\theta_{\rm lim}$) for two different 3D mesh resolutions and two different lens masses. The low and high mass cases are shown in solid and dashed lines, respectively, with their masses represented in terms of their Einstein radiuses $\theta_{\rm E}$. We observe a universal relationship between the accuracy and $|\boldsymbol{\theta}_0|/\theta_{\rm lim}$. HRT maintains sub-percent accuracy beyond $\theta_{\rm lim}$, demonstrating that mesh resolution is the primary determinant of accuracy.

Our fiducial results are obtained using a $512 \times 512 \times 512$ simulation box with a particle spacing of 0.4 Mpc/h and a mesh spacing of 0.4 Mpc/h. We first evolve the particles from initial condition to $a = 1/64$ using 2nd-order Lagrangian perturbation theory. We then simulate the gravitational interaction using the PM algorithm from $a = 1/64$ to today in 64 time steps[6]. From there, we perform ray tracing back to $z_s = 1.034$ across 30 time steps. The lens distirbution is the entire density field in the simulation, and the source distribution is a Dirac delta distribution at z_s. The image plane spans $205' \times 205'$; it includes 1024×1024 pixels with a $0.2'$ pixel

6. This choice differs from Ref. [133]'s, who seeds the initial condition at $a = 1/7$.

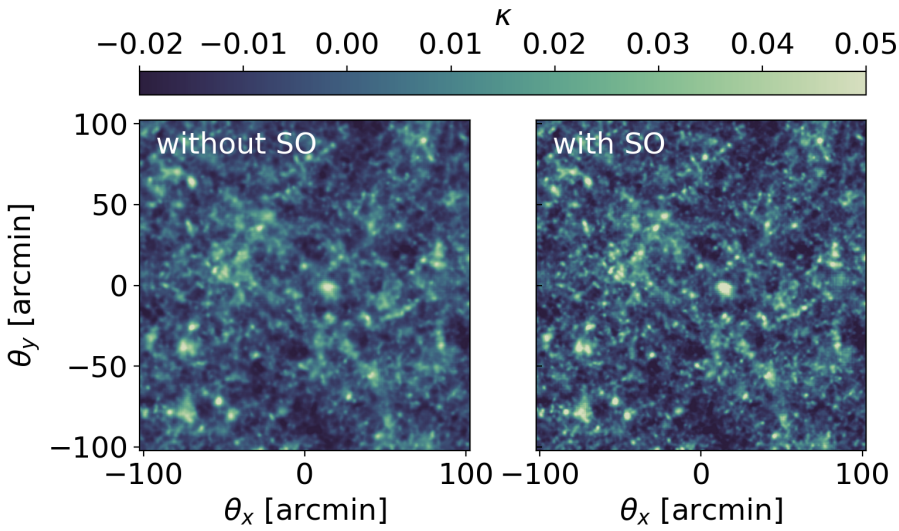

Figure 4.3: Convergence (κ) maps generated using the HRT algorithm on a cosmological simulation. The left panel shows the κ map derived from a density field evolved using the standard PM gravity solver. The right panel illustrates the κ map derived from the same initial cosmological conditions but evolved using the PM algorithm enhanced with the spatial optimization (SO) gravity solver, that sharpens the convergence peaks compared to the regular PM.

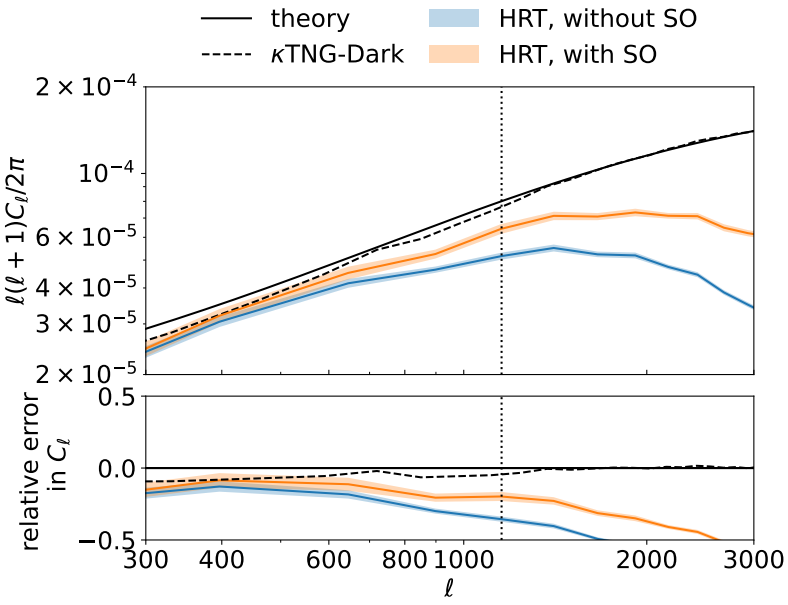

Figure 4.4: The top panel shows the distribution of the convergence power spectra obtained by ray tracing through cosmological simulations using HRT. The bottom panel shows the relative error compared to theory (black, solid) for the same cosmology and $n(z)$. The κTNG results are shown in black, dashed lines for comparison. Simulations where gravity is solved using the standard PM method are depicted in blue, while those using the PM enhanced with the spatial optimization (SO) force-sharpening method are shown in orange. The HRT and the κTNG C_ℓ's are similarly suppressed at low ℓ's because of the finite simulation box size and the lack of large-scale modes. The HRT results align with κTNG at large scales but are suppressed at smaller scales since PM methods cannot resolve interactions at or below the mesh resolution. Enhancing PM with SO mitigates this issue and extends $C_\ell^{\rm HRT}$ to higher ℓ's. For the PM+SO result (where the redshift distribution is a delta function at the same redshift, $z = 1.034$, as the κTNG lensing map), $C_\ell^{\rm HRT}$ agrees with $C_\ell^{\kappa\rm TNG}$ to within 10% for $\ell < 800$ and 15% for $\ell < 1200$, where $\ell = 1200$ is marked by the vertical dotted line.

size. The size and resolution of our simulation box are constrained by the memory capacity of the GPU[7]. Consequently, we do not yet have the hardware capability to conduct ray tracing within a single, monolithic simulation box. Instead, we tile our past light cone by replicating the snapshots 12 times, each modified by a random translation and rotation along the three spatial axes. This tiling strategy, extensively studied in Ref. [181], can produce 10^4 independent realizations of weak lensing power spectra and peak statistics with even a single snapshot. This process has been used in many weak lensing mocks including the κTNG simulations [157]. An example of our ray-traced κ map is shown in the left panel of Fig. 4.3. We run 50 independent simulations and present the distribution of $C_\ell^{\rm HRT}$ along with its comparison to theory (black, solid) and $C_\ell^{\kappa\rm TNG}$ (black, dashed) in the top and bottom panels of Fig. 4.4 (in blue). The plot shows that $C_\ell^{\rm HRT}$ aligns with $C_\ell^{\kappa\rm TNG}$ within 20% for $\ell < 800$ and within 30% for $\ell < 1200$. $C_\ell^{\rm HRT}$ is suppressed on the large scale because our simulation box is small and can only include a limited number of large-scale modes. However, the amount of deficit in power is consistent with the result $C_\ell^{\kappa\rm TNG}$ and the results in obtained by Ref. [133] who performed lensing simulation using the same simulation volume. We also observe $C_\ell^{\rm HRT}$ to be suppressed at high ℓ's. This is because the PM gravity solver cannot accurately resolve gravitational interactions at the mesh scale. This result is comparable to the "DLL without correction" analysis in Ref. [133], who recovers lensing C_ℓ up to $\ell \approx 300$ compared to $C_\ell^{\kappa\rm TNG}$. We recover C_ℓ over a larger ℓ range because of the finer mesh size used in the present work.

Next, we explore the potential of recovering the convergence power spectrum at

7. For this test, we use a H100 NVL GPU on the Vera cluster at the Pittsburgh Supercomputing Center.

higher ℓ ranges by extending the regular PM gravity solver to smaller scales with a spatial optimization (SO) algorithm that sharpens the PM forces near the mesh scale [182]. In Fourier space, the gravitational force is proportional to $(\mathbf{k}/k^2)\delta(\mathbf{k})$. SO modifies the Fourier force kernel by a nonlinear function g:

$$\frac{\mathbf{k}}{k^2} \to \frac{\mathbf{k}}{k^2} g(\mathbf{k}; \boldsymbol{\vartheta}),, \tag{4.33}$$

where $\boldsymbol{\vartheta}$ includes cosmological parameters and simulation configurations. We implement g using symmetry-preserving neural networks, which are trained to align the PM+SO simulations with the GADGET-4 [170] simulations across a wide range of cosmologies and simulation configurations. As illustrated in Fig. 4.5, SO boosts the small-scale $P(k)$ compared to the regular PM algorithm for $k > 1h/\mathrm{Mpc}$, though still slightly lower than the halofit [183] predictions. We run HRT simulations under the same settings as previously discussed but integrate SO with the same initial conditions. An example κ map, shown in the right panel of Fig. 4.3, indicates that SO indeed sharpens the small-scale fluctuations compared to the regular PM simulations. The distribution of power spectra with SO is displayed in Fig. 4.4 in orange. SO consistently improves the convergence power spectrum at all scales, but is especially helpful at high ℓ's. C_ℓ^{HRT} agrees with $C_\ell^{\kappa\mathrm{TNG}}$ to within 10% for $\ell < 800$ and 15% for $\ell < 1200$. In general, the maximum ℓ at which we can recover C_ℓ depends on the 3D mesh resolution l_{3D} and the shape of the lensing kernel. Empirically, we find this scale to be approximately $\ell \approx \frac{\ell_{\mathrm{lim}}}{16} \approx 1200$ for $z_s = 1.034$ when SO is applied (vertical dotted line in Fig. 4.4). We define $\ell_{\mathrm{lim}} = \pi / \left(\frac{\chi(z_s)}{2l_{\mathrm{3D}}}\right)$, since $\frac{\chi(z_s)}{2l_{\mathrm{3D}}}$ is the angular scale of the 3D mesh at the radial comoving distance where the lensing

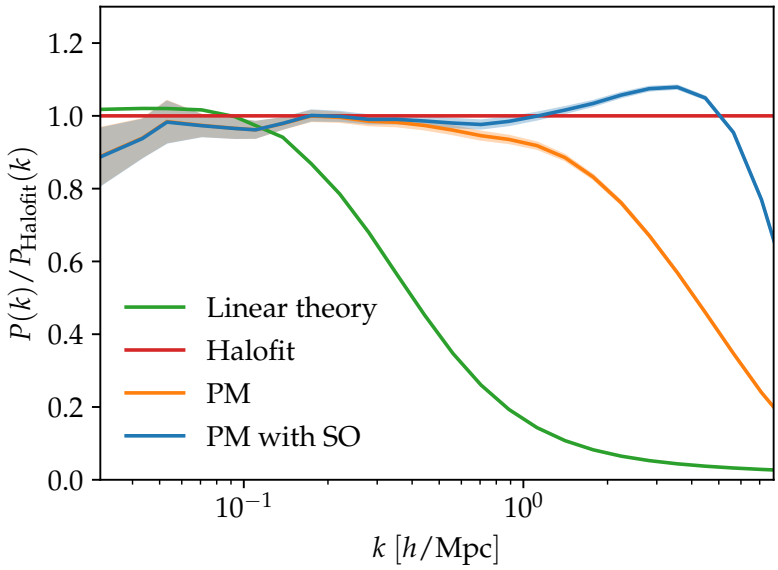

Figure 4.5: Matter power spectrum at $z = 0$ with the regular PM solver (orange) and with PM + spatial optimization (SO, blue). They are compared to the prediction of the Halofit model in red and the linear theory in green. The shaded regions represent the standard error of the mean estimated with 50 independent simulations. The simulation has 512^3 particles with a 0.4 Mpc/h particle spacing. SO corrects for the small-scale interactions between dark matter particles that are ignored by the regular PM, boosting power at high-k.

kernel has the most sensitivity. More generally, when the source distribution is not a delta function, the denominator of ℓ_{lim} should be similarly approximated by the distance to the peak of the lensing kernel.

We anticipate that higher mesh resolution and stronger force-sharpening effects will enable C_ℓ^{HRT} to reach even smaller scales, which is potentially achievable when larger GPU memory becomes accessible in the future. We also conduct convergence tests with varying HRT hyper-parameters in Section 4.9.5.

4.8 Conclusions

We have presented a post-Born, three-dimensional, on-the-fly ray tracing algorithm based on the Hamiltonian dynamics of light rays. This method, termed Hamiltonian ray tracing (HRT), includes the lens-lens coupling and does not assume the Born approximation. Additionally, HRT deflects photons based on the gravitational potential generated by the entire cosmological volume, not just that within a single lens plane. HRT also performs ray tracing on-the-fly. As a result, it integrates well with any gravity solver and computes light ray trajectories and lensing observable maps with minimal computational overhead compared to running the gravity solver alone. We implemented HRT using the `pmwd` library on a GPU platform, demonstrating its accuracy and limitations both for point mass lensing and in generating convergence maps and power spectra for dark matter simulations. For point-mass lensing, HRT yields deflection angles accurate to sub-percent levels above the resolution limit. When applied to cosmological simulations using the standard PM gravity solver with a 0.4 Mpc/h particle spacing, HRT generates lensing convergence maps whose power spectrum aligns with the fiducial κTNG results to within 20% of $C_\ell^{\kappa\mathrm{TNG}}$ for $\ell < 800$ and 30% for $\ell < 1200$. Extending the PM with small-scale force-sharpening method (SO) enables recovery of C_ℓ to within 10% of $C_\ell^{\kappa\mathrm{TNG}}$ for $\ell < 800$ and 15% for $\ell < 1200$. Future works could extend the tests to other scenarios, such as lensing by compensated black holes in an evolving cosmological setting [184] or different weak lensing statistics paired with various redshift distributions.

While HRT should work with any simulation of structure formation, its implementation is particularly easy in frameworks with automatic differentiation capa-

bility, such as `pmwd` which is based on `jax`. The forward mode in automatic differentiation helps to co-evolve the lensing observables, such as cosmic shear, with the light ray deflections, since the Jacobian of the latter involves the former. On the other hand, we would also need the reverse-mode differentiation to compute the likelihood gradient, for example, in FLI applications that involve Hamiltonian Monte Carlo. Automatic reverse-mode differentiation through the whole simulation can be extremely memory consuming, and the adjoint method [126] has been introduced to obtain memory-efficient gradients. The same method can be applied to HRT, which we leave for future development.

The accuracy of HRT is primarily limited by the mesh resolution and the precision of the gravity solver at small comoving scales. If we assume that the smallest scale at which the gravity solver is accurate is $k_{\max} \propto l_{\rm 3D}^{-1}$, and that the smallest angular scale at which HRT is accurate is $\ell_{\max} \propto l_{\rm 3D}^{-1}$, then $\ell_{\max} \propto k_{\max}$. For the current generation of weak lensing surveys like HSC, $\ell_{\max} \approx 1800$ [96], about 2 times higher than the ℓ limit we achieve here. To bridge the gap between simulation and observational data, future work could extend the SO framework to push $k_{\max}$ above 5 h/Mpc where baryonic effects also becomes important. Alternatively, increasing the resolution of the cosmological simulation by two or three-folds, which could be achieved by parallelizing the PM code and the HRT algorithm across multiple GPU devices or nodes, may also prove equally effective.

The HRT algorithm empowers cosmological analysis in several ways. HRT can quickly generate cosmology-dependent ray tracing shear maps, making it suitable for training machine learning models for simulation-based inference. It can help

establishing the connection between cosmology and higher-order statistics in a data-driven manner and aid in estimating cosmology-dependent covariance matrices for summary statistics analyses. As a differentiable ray tracing algorithm, HRT also enables field-level inference that accounts for post-Born effects. While this work focuses on the algorithm and its implementation, future studies could extend this discussion with a detailed analysis of higher-order statistics in the convergence maps simulated by HRT. Additionally, optimizing the differentiation of HRT through the adjoint method could make it more memory-friendly in field-level inference applications.

4.9 Appendix

4.9.1 Hamiltonian dynamics of photons lensed by weak gravitational field

In this section, we derive the EOMs in Eqs. (4.5) to (4.7) from the Hamiltonian principle. Although the procedure is similar to that of Ref. [167], we employ a different metric and present the derivations in full detail. These details, not included in Ref. [167], may prove helpful to some readers.

The general relativistic covariant Hamiltonian for a photon is given by Eq. (4.4) and copied here:

$$H(x^j, p_i, \tau) = c\sqrt{h^{ij}p_i p_j}(1 + 2\frac{\Phi}{c^2}), . \tag{4.34}$$

The photon's EOMs are given by Hamilton's equations

$$\frac{dx^i}{d\tau} = \{x^i, H\} = \frac{\partial H}{\partial p_i},, \tag{4.35}$$

$$\frac{dp_i}{d\tau} = \{p_i, H\} = -\frac{\partial H}{\partial x^i},, \tag{4.36}$$

which we will solve explicitly. The time derivatives of position and momentum are

$$\frac{1}{c}\frac{dx^j}{d\tau} = \frac{\partial\sqrt{h^{kl}p_k p_l}}{\partial p_j}\Big(1+2\frac{\Phi}{c^2}\Big) = \frac{h^{kl}}{2p}\Big(\frac{\partial p_k}{\partial p_j}p_l + \frac{\partial p_l}{\partial p_j}p_k\Big)\Big(1+2\frac{\Phi}{c^2}\Big) = n^j\Big(1+2\frac{\Phi}{c^2}\Big),. \tag{4.37}$$

$$\frac{1}{c}\frac{dp_j}{d\tau} = -2p\frac{\partial_j\Phi}{c^2} - \frac{p_k p_l\partial_j h^{kl}}{2p}\Big(1+2\frac{\Phi}{c^2}\Big) = -p\Big[2\frac{\partial_j\Phi}{c^2} + \frac{n_k n_l\partial_j h^{kl}}{2}\Big(1+2\frac{\Phi}{c^2}\Big)\Big],, \tag{4.38}$$

where the unit momentum vector n^i is

$$n^i \equiv \frac{p^i}{p} = \frac{p^i}{\sqrt{h^{ij}p_i p_j}},. \tag{4.39}$$

n^i is the most important dynamical variable in ray-tracing. To explicitly derive its EOM, we express its time dependence in those of the two independent variables p_i and x^j, and consider the following expansion

$$\frac{1}{c}\frac{dn^i}{d\tau} = \frac{1}{c}\frac{d}{d\tau}\Big(\frac{p^i}{p}\Big) = \frac{\partial(p^i/p)}{\partial p_j}\frac{1}{c}\frac{dp_j}{d\tau} + \frac{\partial(p^i/p)}{\partial x^j}\frac{1}{c}\frac{dx^j}{d\tau},. \tag{4.40}$$

The first term becomes

$$\frac{\partial(p^i/p)}{\partial p_j}\frac{1}{c}\frac{dp_j}{d\tau} = \frac{1}{c}\frac{dp_j}{d\tau}\Big(\frac{h^{ik}}{p}\frac{\partial p_k}{\partial p_j} - \frac{p^i}{p^2}\frac{\partial p}{\partial p_j}\Big) = \frac{1}{c}\frac{dp_j}{d\tau}\frac{h^{ij}-n^in^j}{p}$$
$$= \Big[-2\frac{\partial_j\Phi}{c^2} - \frac{n_kn_l\partial_jh^{kl}}{2}\Big(1+2\frac{\Phi}{c^2}\Big)\Big](h^{ij}-n^in^j),. \tag{4.41}$$

Meanwhile, the second term becomes

$$\frac{\partial(p^i/p)}{\partial x^j}\frac{1}{c}\frac{dx^j}{d\tau} = \frac{1}{c}\frac{dx^j}{d\tau}\Big(\frac{p_k}{p}\partial_jh^{ik} - \frac{p^i}{p^2}\partial_jp\Big) = \frac{1}{c}\frac{dx^j}{d\tau}\Big(n_k\partial_jh^{ik} - \frac{1}{2}n^in_kn_l\partial_jh^{kl}\Big)$$
$$= \Big(n^jn_k\partial_jh^{ik} - \frac{1}{2}n^in^jn_kn_l\partial_jh^{kl}\Big)\Big(1+2\frac{\Phi}{c^2}\Big),. \tag{4.42}$$

Putting the two terms together,

$$\begin{aligned}\frac{1}{c}\frac{dn^i}{d\tau} =& -2\frac{\partial_j\Phi}{c^2}(h^{ij}-n^in^j) - \frac{n_kn_l\partial_jh^{kl}}{2}\Big(1+2\frac{\Phi}{c^2}\Big)(h^{ij}-\cancel{n^in^j})\\ &+\Big(n^jn_k\partial_jh^{ik} - \cancel{\frac{1}{2}n^in^jn_kn_l\partial_jh^{kl}}\Big)\Big(1+2\frac{\Phi}{c^2}\Big)\\ =& -2\frac{\partial_j\Phi}{c^2}(h^{ij}-n^in^j) - \frac{h^{ij}n_kn_l\partial_jh^{kl}}{2}\Big(1+2\frac{\Phi}{c^2}\Big) + n^jn_k\partial_jh^{ik}\Big(1+2\frac{\Phi}{c^2}\Big),.\end{aligned} \tag{4.43}$$

It turns out that we can simplify the last two terms, because

$$\begin{aligned}
-\frac{1}{2}h^{ij}n_k n_l \partial_j h^{kl} + n^j n_k \partial_j h^{ik} &= \frac{1}{2}h^{ij}n^k n^l \partial_j h_{kl} - h^{il}n^j n^k \partial_j h_{kl} \\
&= \frac{1}{2}h^{il}n^j n^k (\partial_l h_{jk} - 2\partial_j h_{kl}) \\
&= -\frac{1}{2}h^{il}n^{(j}n^{k)}(\partial_j h_{kl} + \partial_k h_{jl} - \partial_l h_{jk}) \\
&= -\Gamma^i_{jk} n^j n^k,, \qquad (4.44)
\end{aligned}$$

where we have used $\partial_j h^{ik} = -h^{il}h^{km}\partial_j h_{lm}$ in the first equality, swapped j and l in the (first term of the) second one, symmetrized j and k in the third, and reduced the metric derivatives to the Christoffel symbol Γ^i_{jk} at last. With the result above, we can simplify $dn^i/d\tau$ as follows

$$\frac{1}{c}\frac{dn^i}{d\tau} = -2\frac{\partial_j \Phi}{c^2}(h^{ij} - n^i n^j) - \Gamma^i_{jk} n^j n^k \Big(1 + 2\frac{\Phi}{c^2}\Big),, \qquad (4.45)$$

which agrees with the result in [167].

We are now ready to derive the EOMs of a photon. Until this point, we have not enforced any specific metric on the EOMs. We now select the metric in Eq. (4.3) with coordinates $\mathbf{x} = (\boldsymbol{\theta}, \chi)$. Here, we assume small angle approximations such that $\sin^2(\theta) \approx 1$. The unit momentum vector in this coordinate system is then

$$\mathbf{n} = \Big[\frac{\boldsymbol{v}}{cr}, -1 + \mathcal{O}\Big(\frac{\boldsymbol{v}^2}{c^2}\Big)\Big],, \qquad (4.46)$$

where the first term can be viewed as the normalized peculiar angular velocity on the sky, with $\boldsymbol{v}$ being the 2D transverse peculiar velocity. We shall work in the limits of

weak fields and small angles, which corresponds to first order in Φ/c^2, $\boldsymbol{\nabla}\Phi/c^2$, and $\boldsymbol{v}/c$. The nonzero Christoffel symbols are

$$\Gamma^1_{13} = \Gamma^1_{31} = \Gamma^2_{23} = \Gamma^2_{32} = \frac{r'}{r},, \qquad \Gamma^3_{11} = \Gamma^3_{22} = -rr',, \tag{4.47}$$

where $'$ denotes derivative with respect to χ. We have omitted corrections of $\mathcal{O}(\boldsymbol{\nabla}\Phi/c^2)$ because the Christoffel symbol only appears in the expression $\Gamma^i_{jk} n^j n^k (1 + 2\Phi/c^2)$, and always pairs with at least one $\mathcal{O}(v/c)$ term. Taking time derivatives of n^i for $i = 1, 2$, we obtain

$$\frac{1}{c}\frac{dn^i}{d\tau} = \frac{1}{c^2 r}\frac{dv^i}{d\tau} - \frac{v^i}{c^2 r^2}\frac{dr}{d\tau} \tag{4.48}$$

Consider $i = 1$ of the n^i EOM:

$$\begin{aligned}\frac{1}{c}\frac{dn^1}{d\tau} &= \frac{1}{c^2}(-2h^{11}\partial_1\Phi + 2n^1 n^1 \partial_1\Phi + 2n^1 n^2 \partial_2\Phi + 2n^1 n^3 \partial_3\Phi) - 2\Gamma^1_{13} n^1 n^3 \Big(1 + 2\frac{\Phi}{c^2}\Big) \\ &\simeq -2\frac{\partial_1\Phi}{c^2 r^2} + 2\frac{v^1 r'}{c r^2},.\end{aligned} \tag{4.49}$$

The general case follows similarly by symmetry:

$$\frac{1}{c}\frac{dn^i}{d\tau} = -2\frac{\partial_i\Phi}{c^2 r^2} + 2\frac{v^i}{c}\frac{r'}{r^2},, \tag{4.50}$$

where we ignore the potential and its gradient terms since they are higher order corrections (while more generally $h^{ij} - n^i n^j$ projects in the transverse direction of n^i). Comparing this with the above equation, and approximating $cd\tau \approx -d\chi$ by

ignoring time delay correction of $\mathcal{O}(\Phi/c^2)$ and $\mathcal{O}(v^2/c^2)$, we derive

$$\frac{d\eta^i}{d\chi} \equiv \frac{d(rv^i/c)}{d\chi} = \frac{rv^{i\prime} + r^\prime v^i}{c} = 2\frac{\partial_i \Phi}{c^2} = 2\frac{1}{c^2}\frac{\partial \Phi}{\partial \theta^i},. \tag{4.51}$$

4.9.2 Symplectic integrator

We have purposely written the EOMs in Eqs. (4.5) and (4.6) using $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$, in unit of angle and length, respectively. Either from the Hamiltonian Eq. (4.4) or from the EOMs, we see that the system, under the assumptions above, admits a separable Hamiltonian

$$\tilde{H}(\boldsymbol{\theta}, \boldsymbol{\eta}, \chi) = \underbrace{\frac{\boldsymbol{\eta}^2}{2r^2}}_{T} + \underbrace{2\frac{\Phi(\boldsymbol{\theta}, \chi)}{c^2}}_{V},, \tag{4.52}$$

where $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$ are the canonical coordinate and momentum, χ serves the function of time, and T and V represent kinetic and potential energy. In fact, $\tilde{H}$ also follows from the Fermat principle, and can be derived with Legendre transformation from the Lagrangian in the Fermat action (see [167, 185, 186], especially for the continuous version in the former). This allows us to decouple the effect of the kinetic and the potential energy and rewrite the EOMs in Eqs. (4.5) and (4.6) as

$$\frac{d\boldsymbol{\theta}}{d\chi} = \{\boldsymbol{\theta}, T\},, \qquad \frac{d\boldsymbol{\eta}}{d\chi} = \{\boldsymbol{\eta}, V\},. \tag{4.53}$$

where $\{\}$ is the Poisson bracket.

Formally, we can integrate[8] Eq. (4.53) over the interval $[\chi, \chi + \Delta\chi]$ by

$$\begin{bmatrix} \boldsymbol{\theta} \\ \boldsymbol{\eta} \end{bmatrix} (\chi + \Delta\chi) = e^{(T+V)\Delta\chi} \begin{bmatrix} \boldsymbol{\theta} \\ \boldsymbol{\eta} \end{bmatrix} (\chi), . \tag{4.54}$$

Since the EOMs of $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$ are coupled together, we cannot solve them individually. The KDK integrator utilizes the Baker-Campbell-Hausdorff identity to approximate this evolution operator as the product

$$e^{(H+\tilde{H}_{\text{err}})\Delta\chi} = e^{V\frac{\Delta\chi}{2}} e^{T\Delta\chi} e^{V\frac{\Delta\chi}{2}} \tag{4.55}$$

where $\tilde{H}_{\text{err}}$ denotes the approximation error, which is second order in $\Delta\chi$ [171]. Equation (4.55) shows that we can decompose each $\Delta\chi$ integration at each time step into three consecutive steps, each updating only either the position or the momentum variables. $e^{V\Delta\chi}$ is termed a kick operator since it updates $\boldsymbol{\eta}$ and leaves $\boldsymbol{\theta}$ unchanged, while $e^{T\Delta\chi}$ is called a drift operator since it updates $\boldsymbol{\theta}$ and leaves $\boldsymbol{\eta}$ unchanged.

4.9.3 Adaptive ray mesh

The 2D array of light rays is characterized by its pixel size μ_{2D}, i.e. their spacing at $z = 0$, and the number of pixels $M_{x,y}$. In order to interpolate and transfer the PM forces from the 3D mesh to the rays, we need another intermediate 2D angular mesh of resolution ν_{2D} and size $N_{x,y}$. We call it the ray mesh (and likewise the 3D mesh particle mesh), as explained in Table 4.1.

8. Here, for an operator A and a vector $\boldsymbol{v}$, we use the shorthand $A\boldsymbol{v} = \{\boldsymbol{v}, A\}$.

Generally, $\nu_{2\mathrm{D}} \leq \mu_{2\mathrm{D}}$ and $N_{x,y} > M_{x,y}$. Our adaptive ray mesh needs to address two problems: 1) we need to vary the 2D mesh resolution $\nu_{2\mathrm{D}}$ as a function of χ (Section 4.6.3) and 2) we need to add padding to the ray mesh. Padding is crucial for accurate smoothing and deconvolution in the computation of the kick operator, as it alleviates the effects of periodic boundary assumptions inherent to FFT.

A 3D mesh cell at χ corresponds to an angular size $l_{3\mathrm{D}}/r(\chi)$. Therefore, for a lens plane that spans $[\chi_a, \chi_b]$, we define the angular resolution limit λ_{lim} as in Eq. (4.30):

$$\lambda_{\mathrm{lim}} = \max\Big(\frac{2, l_{3\mathrm{D}}}{r_a + r_b}, \mu_{2\mathrm{D}}\Big), , \tag{4.56}$$

in which the former dominates near the observer as limited by the PM force resolution, and the latter takes effect at early times. Let us fix the ray mesh spacing as

$$\nu_{2\mathrm{D}} = \iota \lambda_{\mathrm{lim}}, , \tag{4.57}$$

where $0 < \iota \leq 1$ is an accuracy parameter. The smaller ι is, the finer the ray mesh. Take the x-direction for example, we want

$$N_x \nu_{2\mathrm{D}} \geq M_x \mu_{2\mathrm{D}} + p_{\mathrm{min}} \lambda_{\mathrm{lim}}, , \tag{4.58}$$

where p_{min} is the minimum padding width in unit of λ_{lim}. The above conditions reduce to

$$N_x = \Big\lceil M_x \frac{\mu_{2\mathrm{D}}}{\nu_{2\mathrm{D}}} + \frac{p_{\mathrm{min}}}{\iota} \Big\rceil_{\mathrm{FFT}}, . \tag{4.59}$$

where we round up the mesh size to an integer that is efficient for FFT, e.g. powers

of 2, while also saving the number of compilations that can affect `jax` performance. The y-dimension follows accordingly.

4.9.4 Other considerations and convergence tests for point mass lensing

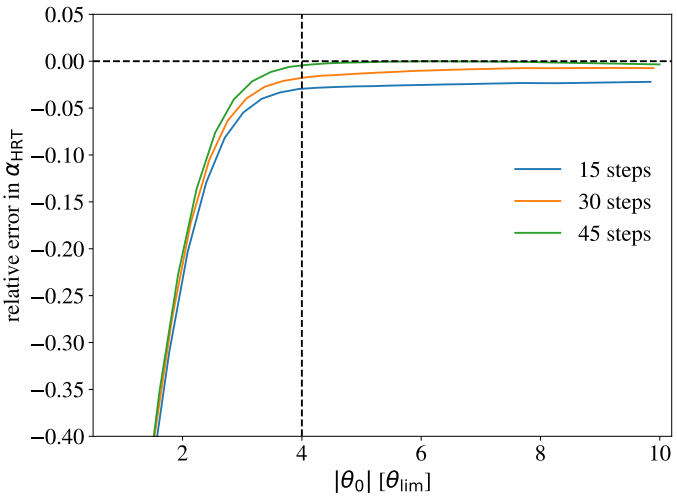

Figure 4.6: Relative error of $\boldsymbol{\alpha}_{\mathrm{HRT}}$ versus θ_{lim} as a function of the number of HRT integration steps used for a point mass lensing task. This figure is to be compared to the blue line in Fig. 4.2 which has the same experimental setup with 45 integration steps.

In the context of point mass lensing, we explore how the accuracy of HRT is influenced by the boundary conditions of the 3D mesh, the number of integration steps, the 2D ray mesh accuracy parameter ι, and the 2D ray mesh padding size p_{min} (Section 4.9.3).

First, when we perform ray-tracing on a point mass lens, we compute the gravitational potential field using FFT with periodic boundary conditions. This is an

approximation of the gravity solver and not the ray-tracing algorithm itself. To disentangle the error induced by the gravity solver from those due to HRT, we have compared $\boldsymbol{\alpha}_{\mathrm{HRT}}$ against a theoretical result that incorporates periodic boundary conditions throughout this work. To model the effect of the periodic boundary condition, we assume there is not only one but also infinitely many periodic images of the point mass lens, at comoving positions $\mathbf{m} \in \{(0,0), (\pm L_x, 0), (0, \pm L_y), \cdots\}$, where $L_{x,y}$ are the side lengths of the simulation box. The leading order deflection in Eq. (4.32) is replaced by its periodic summation:

$$\hat{\alpha}\frac{\boldsymbol{\theta}_0}{|\boldsymbol{\theta}_0|} = \frac{4GM}{c^2}\frac{\mathbf{b}}{b^2} \longrightarrow \frac{4GM}{c^2}\sum_{\mathbf{m}}\frac{\mathbf{b}-\mathbf{m}}{|\mathbf{b}-\mathbf{m}|^2},, \tag{4.60}$$

where $\mathbf{b} = (\theta_{0x}\chi_{\mathrm{l}},\ \theta_{0y}\chi_{\mathrm{l}})$ is the impact parameter of the photons. Equation (4.60) reduces to Eq. (4.32) when only $\mathbf{m} = (0,0)$ is considered. Throughout this work, we account for the nearest 32^2 periodic images of the lens mass when computing $\boldsymbol{\alpha}_{\mathrm{theory}}$. The theoretical result converges within numerical precision.

The numerical experiments in Section 4.7.1 use 45 integration steps and assume a padding of $\iota = 0.5$ and $p_{\mathrm{min}} = 256$. We utilize the same setup as in Section 4.7.1 to evaluate how the accuracy of HRT depends on these hyper-parameters. Figure 4.6 illustrates that $\boldsymbol{\alpha}_{\mathrm{HRT}}$ converges to the theoretical values as the number of integration steps increases. With too few integration steps, $\boldsymbol{\alpha}_{\mathrm{HRT}}$ is typically lower than the truth. The two panels in Fig. 4.7 demonstrate the effect of the ray mesh hyper-parameters: the accuracy parameter ι and the padding size p_{min}. The left panel uses ($\iota = 1,\ p_{\mathrm{min}} = 4$) and the right panel uses ($\iota = 1,\ p_{\mathrm{min}} = 256$). These are

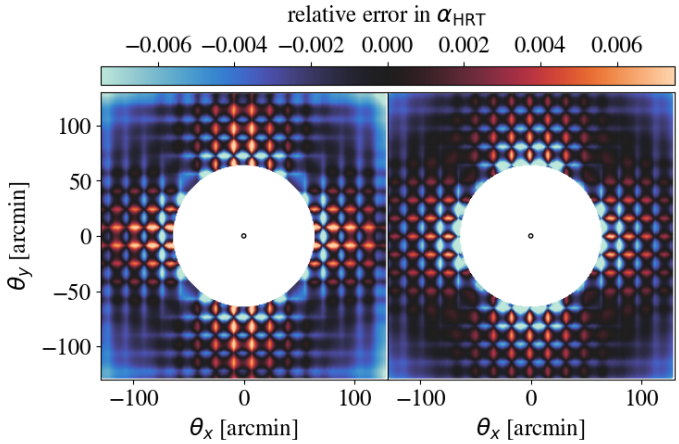

Figure 4.7: The effect of ray mesh hyper-parameters: the accuracy parameter ι and the padding size p_{min}. The two panels show $\boldsymbol{\alpha}_{\mathrm{HRT}}$'s residual errors on the image plane for the point mass test. The left panel uses ($\iota = 1,\ p_{\mathrm{min}} = 4$) and the right panel uses ($\iota = 1,\ p_{\mathrm{min}} = 256$). They are compared to panel (b) of Fig. 4.1 which uses ($\iota = 0.5,\ p_{\mathrm{min}} = 256$).

compared to panel (b) of Fig. 4.1, which has the same experimental setup but with ($\iota = 0.5,\ p_{\mathrm{min}} = 256$). In both cases, HRT achieves sub-percent accuracy, showing that increased padding improves $\boldsymbol{\alpha}_{\mathrm{HRT}}$ accuracy near the boundary and higher ι results in more prominent discretization features of the particle and ray meshes. The result also shows that, in practice, we do not need to carry a p_{min} that is as large.

4.9.5 Convergence tests for weak lensing power spectrum

We compute the convergence power spectrum by ray tracing to a source plane at $z_{\mathrm{s}} = 1.034$. We tile the past light cone with simulation boxes of a 512^3 mesh and 512^3 particles with a particle/mesh spacing of 0.4 Mpc/h, similar to Section 4.7.2. We vary the number of ray tracing time steps and the ray mesh accuracy parameter ι, as shown in Fig. 4.8, and find that the power spectrum is not sensitive to these

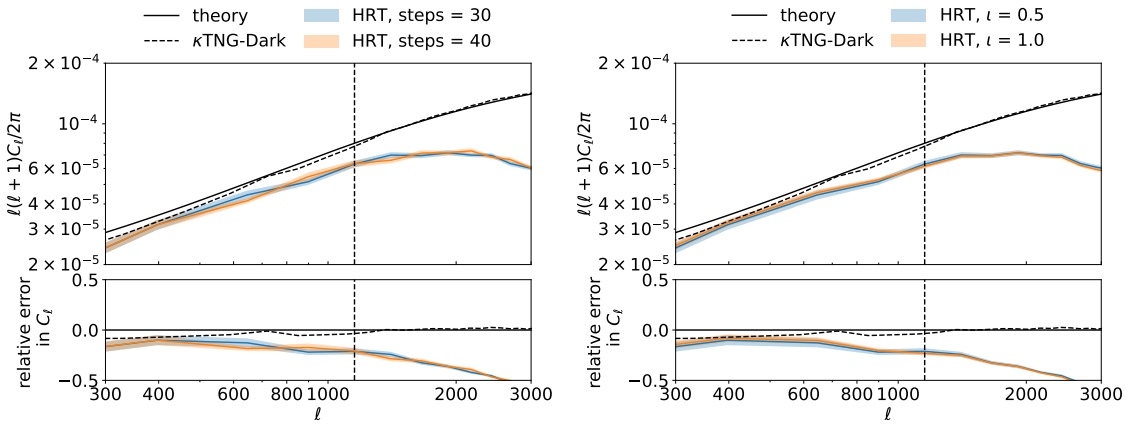

Figure 4.8: Relative error of the SO-enhanced C_ℓ as a function of the number of HRT integration steps (left) and of the HRT ray mesh accuracy parameter ι (right). The shaded regions show the standard error of the mean over 25 independent simulations. Similar to Fig. 4.4, the black solid line shows the theory prediction, the dashed line shows the κTNG simulation, and the vertical dotted line shows the $\ell = 1200$ scale cut.

hyper-parameters.

CHAPTER 5

MAP-LEVEL BARYONIFICATION: UNIFIED TREATMENT OF WEAK LENSING TWO-POINT AND HIGHER-ORDER STATISTICS

5.1 Abstract

Precision cosmology benefits from extracting maximal information from cosmic structures, motivating the use of higher-order statistics (HOS) at small spatial scales. However, predicting how baryonic processes modify matter statistics at these scales has been challenging. The baryonic correction model (BCM) addresses this by modifying dark-matter-only simulations to mimic baryonic effects, providing a flexible, simulation-based framework for predicting both two-point and HOS. We show that a 3-parameter version of the BCM can jointly fit weak lensing maps' two-point statistics, wavelet phase harmonics coefficients, scattering coefficients, and the third and fourth moments to within 2% accuracy across all scales $\ell < 2000$ and tomographic bins for a DES-Y3-like redshift distribution ($z \lesssim 2$), using the FLAMINGO simulations. These results demonstrate the viability of BCM-assisted, simulation-based weak lensing inference of two-point and HOS, paving the way for robust cosmological constraints that fully exploit non-Gaussian information on small spatial scales.

5.2 Introduction

Weak gravitational lensing directly probes the total matter distribution in the Universe, making it a powerful cosmological tool. Unlike galaxy clustering, which is biased by galaxy-matter relationships, weak lensing is sensitive to all matter, both dark and baryonic. Measurements of two-point weak lensing statistics have provided some of the tightest constraints on cosmological parameters, particularly $S_8 = \sigma_8\sqrt{\Omega_m/0.3}$, which captures the degeneracy between the amplitude of matter fluctuations (σ_8) and matter density (Ω_m) [95, 96, 187–191].

There is growing interest in extracting non-Gaussian features from the lensing map, both at the field level [18, 26, 45, 87, 89, 192] and via higher-order statistics (HOS). In particular, many HOS have been successfully applied to data have shown significant gains in constraining power (e.g., Gatti *et al.* 15, Fluri *et al.* 193, Gatti *et al.* 194, Jeffrey *et al.* 195, Marques *et al.* 196, Novaes *et al.* 197, Cheng *et al.* 198). They are also expected to improve our constraining power on extended cosmological models, such as those with primordial non-Gaussianities [199] or modified theories of gravity [200, 201]. Unlike two-point statistics, however, most HOS lack analytical predictions and rely on realistic weak lensing simulations for modeling.

Weak lensing observables – both standard two-point and HOS – probe low redshifts ($z < 1$) and small physical scales ($r < 10\,\mathrm{Mpc}$), making them particularly sensitive to non-linear structure formation and astrophysical processes such as supernova and active galactic nucleus feedback, collectively referred to as baryonic feedback. These astrophysical processes alter both the distribution and thermodynamics of baryons, and affect the dark matter distribution through gravitational

coupling (e.g., Gnedin *et al.* 202, Duffy *et al.* 203, Van Daalen *et al.* 204, Semboloni *et al.* 205, Anbajagane *et al.* 206, Shao *et al.* 207, McCarthy *et al.* 208, Broxterman *et al.* 209, Schaller *et al.* 210, Schaller and Schaye 211). These effects are important on small and intermediate scales, extending into the quasi-linear regime ($\lesssim 5\,\mathrm{Mpc}$), where they contribute significantly to the lensing signal. If unaccounted for, baryonic effects can bias cosmological inferences [194, 212–215]. For example, the observed discrepancy in S_8 between weak lensing and CMB measurements could arise from such mis-modeling rather than a fundamental deviation from the standard cosmological model [e.g., 214]. While a very conservative approach is to remove scales affected by baryons or to implement projection schemes that mitigate their impact [187, 188, 216, 217], accurate modeling remains the preferred solution.

On the modeling side, the baryonic imprints on the matter density field are typically studied using high-resolution cosmological hydrodynamical simulations [see 218, for a review]. However, there is significant freedom in defining the "subgrid'' physics in these simulations – processes like star formation and collisional cooling that occur below the simulation' s resolution scale. Different but *equally well-motivated* subgrid prescriptions result in a wide range of predictions for the distribution of galaxy properties [e.g., 206, 207, 219–227], and consequently for the baryonic imprints on the non-linear matter power spectrum [e.g., 204, 211, 214, 228, 229]. While it is computationally unfeasible to explore broad cosmological and baryonic parameter spaces with these simulations,[1] they are nonetheless essential for validating and calibrating

1. Simulation suites like CAMELS [230] provide a wide variety of simulated predictions spanning this broad parameter space. However, these simulations are performed in cosmologically small volumes ($V \sim (25\,\mathrm{Mpc}/h)^3$) and so cannot be reliably used to explore the impact of baryons on statistics summarizing the large-scale matter distribution.

simpler, more practical parametric models of baryonic effects.

Several parametric models have been developed to flexibly describe the impact of baryons on weak lensing observables. A well-known example is HMCode2020 [231], an augmented halo model designed to accurately predict the non-linear matter power spectrum across a wide range of cosmologies. To account for baryonic feedback, HMCode2020 modifies the concentration–mass relation through a parametric model that accounts for mass and redshift-dependence and is fitted to the matter power spectra of the BAHAMAS simulations [232]. More recent examples include the power spectrum emulator trained on the FLAMINGO simulations [210] and a wavenumber-dependent rescaling of the difference between the non-linear and linear gravity-only power spectra [211, 214]. Another successful approach is the baryonic correction model (BCM), a phenomenological method that mimics baryonic effects by perturbing the density field of dark matter-only (DMO) simulations via post-processing [3, 233, 234]. These modified outputs can then be used to build emulators of the matter power spectrum that include such baryonic imprints. The BCM has demonstrated sufficient flexibility to reproduce a wide range of feedback scenarios seen in hydrodynamical simulations [234–236], and has already been applied to analyze two-point statistics from wide-field surveys [237–239].

Most of the recipes described above have been extensively tested in the context of two-point statistics. In principle, the BCM approach – via its use of N-body simulations – allows the construction of emulators for any observable, including HOS. However, the robustness of the model predictions for beyond two-point functions, in the context of analyzing current and future weak lensing datasets, remains largely

unproven. Some studies have begun to explore this direction: [205] proposed a halo-based model which corrects the baryonic effects on the two and three-point functions. [240] showed that the BCM can simultaneously fit the matter power spectrum and bispectrum of hydrodynamical simulations for snapshots at different redshifts. [241] compared weak lensing peak counts from the full hydrodynamical simulation IllustrisTNG to those from a baryon-corrected version of its dark matter-only (DMO) counterpart, IllustrisTNG-Dark. They found that the BCM provided an insufficient fit when the noise levels match those expected from future surveys like LSST and Euclid. However, their analysis relied on BCM parameters calibrated solely to the power spectrum, whereas a joint fit that included peak counts could have improved performance. Moreover, they did not consider other types of HOS. More recently, Anbajagane *et al.* [3] performed a joint fit of second-, third-, and fourth-moments of the matter density and gas pressure fields measured in IllustrisTNG, finding statistical agreement with the model on scales above $\sim 1\,\mathrm{Mpc}$ and across multiple redshifts. However, this test was also limited to projected slices at fixed redshifts and did not validate against realistic weak lensing observables.

We show that the BCM is sufficiently flexible for simulation-based analyses by applying it to the FLAMINGO simulations, a state-of-the-art suite that features multiple baryonic feedback models, matching DMO runs, and is calibrated against observed $z = 0$ galaxy stellar mass function and the galaxy cluster gas fractions [242, 243]. We baryonify the DMO simulations, generate realistic weak lensing maps, and build an emulator for both two-point and higher-order statistics (HOS) —including scattering transform coefficients [244], wavelet phase harmonics [245], and third- and

fourth-order moments. The emulator accurately recovers the summary statistics of the hydrodynamical runs, and we show that appropriate BCM parameter choices can jointly fit all statistics across the FLAMINGO feedback models down to $\ell \approx 2000$.

The paper is organized as follows. § 5.3 discusses the BCM model. § 3.4 describes the FLAMINGO simulations and the DES Y3-like tomographic weak lensing convergence maps. § 5.5 defines the relevant HOS and how they are measured in simulations. § 5.6 describes the procedure for fitting the BCM model to the FLAMINGO feedback variants, and § 5.7 presents the results. We conclude in § 5.8. Supporting figures and tables are provided in the Appendix.

5.3 Baryonification of simulations with BaryonForge

The BCM introduced in Schneider and Teyssier [233] is a phenomenological framework that does not require specifying any small-scale subgrid physics. The method is parameterized by pairs of halo density profiles: one profile representing a halo in a DMO simulation, and another representing the halo in a hydrodynamical simulation (DMB). The latter includes contributions from dark matter (DM) and baryons (B), in the form of stars, gas, satellite galaxies, and dark matter. There exist different models for the exact form and parameterization of these profile components [4, 221, 233–235]. Here we use the model of Giri and Schneider [236], which is a slightly modified version of the model in Schneider *et al.* [234]. We refer the reader to those works for detailed discussions. Briefly, the total matter in the DMO case follows a Navarro–Frenk–White (NFW) profile [246, 247] with a power-law cut-off at large scales, $r > 4R_{200\mathrm{c}}$ [248]. The stellar density follows an exponential profile,

while the gas density follows a modified generalized NFW (gNFW) profile. The collisionless matter (dark matter and satellite galaxies) is modeled as a modified NFW profile that accounts for adiabatic relaxation due to the presence of the star and gas components. In our implementation, each halo is assigned a NFW concentration using the semi-analytic model of Diemer and Kravtsov [249].

The DMO and DMB profiles can be expressed via the quantity $M(r)$, the mass enclosed within a radius r, which is a monotonically increasing function. The DMO and DMB enclosed mass profiles are then combined to compute a displacement function,

$$\Delta d(r) = M_{\rm DMB}^{-1}\left(M_{\rm DMO}(r)\right) - r. \tag{5.1}$$

where the functions above are implicitly also functions of the halo mass, $M_{\rm 200c}$. Here, the function $M_{\rm DMB}^{-1}$ takes a mass as its argument, and returns the radius that encloses that mass under the DMB model. This displacement function represents the radial offset required to transform a DMO matter distribution into a DMB one.

The BCM proceeds as follows. We start with a DMO simulation with a halo catalog. For each halo above a the mass threshold $10^{13} M_\odot/h^2$, we identify all (dark matter) particles within a predetermined radius. Then, given the halo mass $M_{\rm 200c}$ and the radial separation of the particles from the halo, we evaluate the displacement $\Delta d(r)$. Next, we radially[3] offset the particles by $\Delta d(r)$ from the halo center. The

2. We chose a threshold of $10^{13}\, M_\odot/h$ for computational reasons: the halo mass function is dominated by low-mass halos, which would exponentially increase the pipeline's runtime. The impact of excluding these small halos is discussed in the results section.

3. Anbajagane *et al.* [3] found the impact of anisotropic (elliptical) offsets on the moments of the density field to be sub-percent (see their Figure 7).

displacements are generally positive on the scale of the halo radius, as the dark matter particles must be moved outwards to mimic the impact of extended gas distributions, and are generally negative on scales a hundredth of the halo radius, where the dark matter particles must coalesce to mimic the centrally peaked stellar distribution. The final particle distribution represents a density field that mimics the impact of baryonic processes.

One can then vary the parameters controlling the mass profiles and thereby obtain different corrections to the simulated density field. The profiles specified by a given model can then also be used to predict other quantities, such as the integrated stellar/mass as a function of halo mass. The BCM approach can simultaneously fit variations in the matter power spectrum alongside other quantities such as the scaling relations of gas and star fractions with halo mass, the halo gas profiles, etc. [234, 236, 239, 250].

Our discussions thus far describe the application of BCM to three-dimensional particle snapshots. However, for the purpose of weak lensing, it is computationally expensive to apply the BCM to each individual snapshot. To produce accurate lensing convergence maps, we require fine redshift resolution, e.g., $\mathcal{O}(100)$ snapshots, which increases the computational load of this task. The snapshot-based approach is also particularly expensive for the purpose of simulation-based inference, where of order $\mathcal{O}(10^4)$ simulations are needed to infer cosmological constraints. Therefore, we use the method outlined in Anbajagane *et al.* [3] which circumvents the computational cost by applying the BCM directly on 2D density shells on the full-sky.[4] This

4. An earlier version of this approach is also discussed in Fluri *et al.* [251], but does not contain a number of modeling elements introduced in Anbajagane *et al.* [3].

is done by rewriting the model in terms of *projected* profiles and distances, rather than three-dimensional ones. Instead of particles, our units are the 2D pixels in the full-sky maps. For each pixel, we apply displacements from all halos — where the given pixel is contained with an aperture of $20R_{200c}$ around each selected halo — and accumulate the contributions from all halos to compute the total displacement of the pixel. This extension allows for efficient modeling of higher-order correlations in weak lensing observables by making the BCM more scalable, and therefore makes the method more easily applicable to large datasets from current and upcoming cosmological surveys.

All baryonification in this work is carried out using the publicly available BaryonForge[5] package [3]. We direct readers to Anbajagane *et al.* [3] for more details on the model and method choices. For reference, generating 12 different baryonified simulations, each simulation consisting of 60 tomographic density field shells with `NSIDE` = 2048, takes approximately 4 hours on an AMD EPYC 7763 CPU node with 128 cores and 512 GB of memory. However, we only use a fraction of the memory as the memory footprint of the procedure is set by the density shells and the halo catalogs. For this work, the total memory footprint is less than 10 GB across all shells.

The BCM parameters are listed in Table 5.1, along with their ranges and descriptions. As we will see, the three most important parameters – the ones that can reproduce most of the variation across the hydrodynamic variants – are:

- M_c: a mass scale at which the characteristic slope of the gas profile is $\beta = 1.5$.

5. `https://github.com/DhayaaAnbajagane/BaryonForge`

Halos above/below this mass-scale have more/less cored gas profiles.

- $\theta_{\rm ej}$: a physical scale quantifying how extended the gas distribution is, expressed in units of R_{200c} of the halo. We will see that this ratio is of order 5, so the gas has quite an extended distribution.

- η: The power-law scaling of the stellar mass fraction with halo mass, $M_{\rm star}/M_{\rm 200c} \propto M_{\rm 200c}^{\eta}$.

The exact parameters we vary are primarily informed by the analysis of Giri and Schneider [236]. Compared to their work we also add the redshift scaling parameters ν_X (though, we will show below these are not required for accurate predictions) and η. The latter is added to give the model more flexibility in varying the relative fractions of the gas and star components in a halo.

5.4 Simulations

5.4.1 FLAMINGO simulations

Our analysis relies on the FLAMINGO suite of cosmological hydrodynamical simulations, all of which share the initial conditions of a DMO simulation. We provide a brief summary here; for more details, see [242, 243]. The simulations were run using the SPHENIX smoothed particle hydrodynamics implementation [252] within SWIFT [253]. FLAMINGO includes radiative cooling and heating [254], star formation [255], time-dependent stellar mass loss [256], massive neutrinos [257], and kinetic supernova and stellar feedback [258, 259]. The simulations are calibrated to match

Param	Prior	Description
ν_{M_c}	[-5, 5]	The redshift scaling of the gas mass scale.
$\nu_{\theta_{ej}}$	[-5, 5]	The redshift scaling of the ejection radius.
γ	[0.1, 8]	The slope of the generalized NFW profile at $r \sim R_{ej}$.
$\log_{10}\theta_{co}$	[-4.0, -0.1]	The logarithm of the radial scale of the collapsed gas, $R_{ej} = \theta_{co} R_{200c}$.
$\log_{10}\eta_\delta$	[-3.0, 0.0]	The power-law scaling for the central galaxy stellar fraction–halo mass relation.
δ	[2, 20]	The slope of the generalized NFW profile at $r \gg R_{ej}$.
μ_β	[-3, 3]	The mass-dependent scaling of β, which controls the slope between $R_{co} < r < R_{ej}$.
$\log_{10}\frac{M_c}{M_\odot}$	[12.0, 16.0]	The logarithm of the mass-scale where the characteristic gas slope transitions.
$\log_{10}\eta$	[-3.0, 0.0]	The logarithm of the power-law scaling of stellar fraction–halo mass relation.
θ_{ej}	[0.8, 20]	The radial scale of the ejected gas, $R_{ej} = \theta_{ej} R_{200c}$.

Table 5.1: The input parameters to the profiles of the BCM. For each parameter, we list its prior range and a description of the parameter's physical meaning. See Anbajagane *et al.* [3] for the specific equations that these parameters govern. The prior ranges are similar to those in [3, 4] and are used in § 5.6.1 to construct samples of the baryon-corrected maps and to train the BCM emulator.

the galaxy stellar mass function at $z = 0$ and the gas fraction of low-z clusters and groups of clusters inferred from X-Ray and weak lensing observations [243]. Halos are identified using a three-dimensional friends-of-friends (FoF) algorithm and we utilize the corresponding FoF masses of these halos.[6]

We focus on AGN feedback, implemented in two forms: thermal [260], where particles in the AGN vicinity are heated, and kinetic jet-like, where particles are given momentum kicks along the black hole (BH) spin axis [261]. We consider nine variations in AGN feedback strength from the models available in the FLAMINGO simulations, all in 1 Gpc^3 volumes:

- **Fiducial**: This model features thermal, AGN feedback without jets. Outflows follow paths of least resistance, forming buoyant high-entropy gas bubbles in clusters. Energy is injected into the nearest gas particle, reducing feedback at large distances compared to mass-weighted schemes. The AGN heating temperature, ΔT_{AGN}, is calibrated to match observations such as the cluster gas fractions. This model is labeled Fiducial (L1_m9) per FLAMINGO convention.

- Thermal AGN –cluster gas and galaxy stellar mass function variations: We include **6 variations** of the fiducial model where AGN feedback remains thermal, but the simulations feature different cluster gas fractions (Δf_{gas}) and/or

6. While it is more correct to use spherical overdensity (SO) masses for the baryonification method, we did not have the relevant SO masses for every simulation of interest during the start of this project when the baryonified maps were generated. We have thus used FoF masses instead, which were available for all simulations. While the use of FoF masses over SO masses will lead to some difference (and cause small shifts in the inferred parameters), it does not affect our ability to quantify the flexibility of the baryonification model for the statistics of interest.

galaxy stellar mass functions ($\Delta\Phi_*$), shifted by a specified number of observational standard deviations (σ). These variations lead to different AGN feedback strengths. The models are labeled fgas+2σ, fgas−2σ, fgas−4σ, fgas−8σ, M−σ, and M−σ_fgas−4σ.

- Jet-like AGN Feedback: The **two additional models**, Jet and Jet_fgas−4σ, replace the Fiducial thermal AGN feedback with kinetic, jet-like feedback. Compared to the Fiducial model, Jet has slightly weaker stellar feedback, while Jet_fgas−4σ features higher jet velocity.

We apply the baryonification procedure described in § 5.3 to the full-sky spherical overdensity shells of the DMO simulation to generate baryonified realizations. See Appendix A of Schaye *et al.* [242] for details on how the lightcone is generated. Our procedure for making lensing maps is detailed below in Section 5.4.2. All simulations – both hydrodynamical and baryonified – adopt the best-fit ΛCDM cosmology from DES Y3 "3×2pt + All Ext." [10], which is $h = 0.681, \Omega_{\rm m} = 0.306, \sigma_8 = 0.807$. For the remaining parameters, see Table 4 in Schaye *et al.* [242].

Finally, to investigate the cosmology dependence of baryonic feedback in § 5.7, we also consider simulations run at the "LS8" cosmology [242, 262], which has a lower amplitude of the power spectrum: $S_8 = 0.766$ compared to 0.815 for the fiducial cosmology. The simulations include DMO (LS8_DMO), the fiducial AGN (LS8), and the strong AGN (LS8_fgas-8σ) variants. We did not apply the BCM to the LS8_DMO simulation and used these three simulations exclusively for the cosmology dependence analysis in § 5.7.

5.4.2 DES-like weak lensing convergence maps

We post-process the FLAMINGO simulations to generate weak lensing convergence maps, κ. Each FLAMINGO simulation provides lightcone outputs in the form of HEALPix maps [54] as a function of redshift. The native HEALPix resolution of `NSIDE` = 16384 is degraded to `NSIDE` = 2048 for this project (corresponding to a pixel scale of approximately 1.4 arcminutes), with shell thickness $\Delta z = 0.05$. We produce noiseless, full-sky convergence maps assuming the DES Y3 weak lensing redshift distributions in Fig.5.1 from Gatti *et al.* 263, Myles *et al.* 264. Maps are made assuming the Born approximation.[7] Mimicking the DES Y3 tomographic sample, we generate 4 convergence maps by summing the shells weighted by the appropriate redshift distribution, enabling us to study baryonic effects as a function of redshift. Fig.5.1 shows the resulting 2D DMO power spectra in the four bins compared to theoretical predictions from the Euclid Emulator[2], with agreement at the 1–2% level, consistent with the emulator' s accuracy. We do not add the signal from the intrinsic alignments (IA) of galaxies to our maps, and the coupling of IA and baryonic feedback is minimal [266, 267].

Since our focus is on the relative impact of baryonic feedback compared to the DMO simulation, we do not forward-model the DES Y3 noise or mask, and work exclusively with full-sky, noiseless convergence maps. In total, we use three sets of

7. Petri *et al.* [158] show that post-born corrections, estimated via a full ray-tracing algorithm, result in of $\mathcal{O}(10\%)$ corrections to statistics on small-scales. This is a statement on the absolute accuracy of a simulation compared to the observed lensing process. In this work we are focused on relative differences between simulations, where the relevance of this effect is greatly suppressed. Note also that Broxterman *et al.* [265, see their Figure 1] show the Born approximation produces a consistent (within 1%) result for the angular power spectrum of the matter field for scales up to $\ell < 10^4$

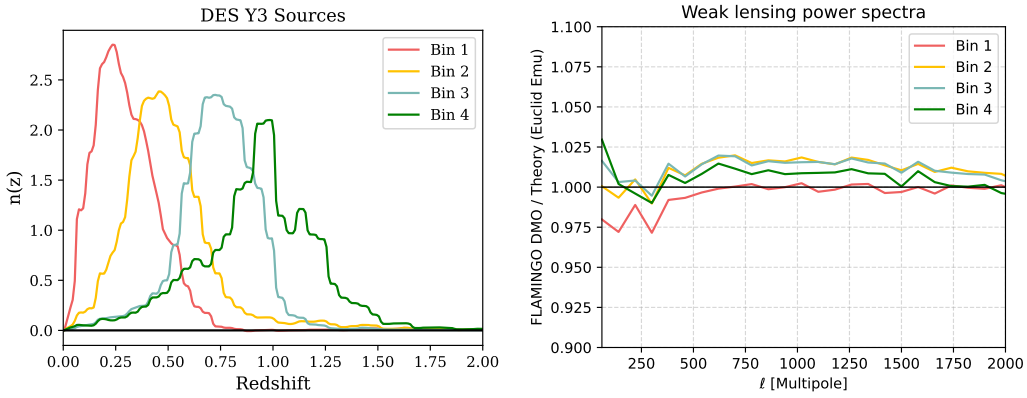

Figure 5.1: *Right:* redshift distribution of the four tomographic bins of the DES Y3 wak lensing sample considered here. *Left:* Power spectra of the 4 noiseless, full-sky convergence maps produced from the FLAMINGO DMO simulation, using the DES Y3 redshift distributions, compared to predictions from EuclidEmu [2].

tomographic maps:

- $\kappa_{\mathbf{dmo}}$: one set from the DMO simulation,
- $\kappa_{\mathbf{flamingo}}$: 9 sets from the nine FLAMINGO feedback variants,
- $\kappa_{\mathbf{bary}}$: $\sim 5,000$ sets, generated by baryonifying $\kappa_{\mathbf{dmo}}$ following the BCM prescription in § 5.3, discussed further in § 5.6.1.

5.5 Statistics

We now describe the theory, measurement, and scale cuts of the statistics used here: the power spectra ($C(\ell)$), the scattering transform coefficients (ST), the wavelet phase harmonic coefficients (WPH), and the third- (m_3) and the fourth-moments (m_4) of the κ maps. As an example, the statistics of $\kappa_{\mathbf{flamingo}}$ are shown in Fig. 5.2 in colored lines. We also summarize all the statistics in Tab. 5.2 at the end of the

section. All statistics are computed on maps of $N_{\rm side} = 1024$, corresponding to a pixel scale of $3.44'$. At $z = 0.5$, which is where the lensing kernel peaks [e.g., 188], this corresponds to a comoving distance of $d = 1.9$Mpc. This is computed assuming the cosmology of the Flamingo simulation used in this work.

5.5.1 Power Spectrum

The power spectrum of a κ map is defined as

$$C(\ell) = \frac{1}{2l+1} \sum_m |\hat{\kappa}_{\ell m}|^2 \,, \tag{5.2}$$

where $\hat{\kappa}_{\ell m}$ is the spherical harmonic transform of κ that is downgraded to $\texttt{NSIDE} = 1024$. We consider the auto-power spectra of the 4 DES redshift bins, linearly binned in $\ell \in [20, 2020]$.

A main objective is to validate BCM's weak lensing statistics to spatial scales smaller than the previous DMO analyses. For example, the DES Y3 weak lensing simulation-based inference with HOS was limited to $\ell_{\rm max} = 1000$ [15, 268], because the analysis assumed a DMO model and could not model smaller scales contaminated by baryonic feedback. Here, we aim to validate the BCM to $\ell_{\rm max} = 2000$. We do not attempt to model $\ell > 2,000$ due to 1) increased systematic effects and a low signal-to-noise ratio in real data, and 2) the computational challenge of producing enough simulations with low shot noise at high ℓ in the context of simulation-based inference.

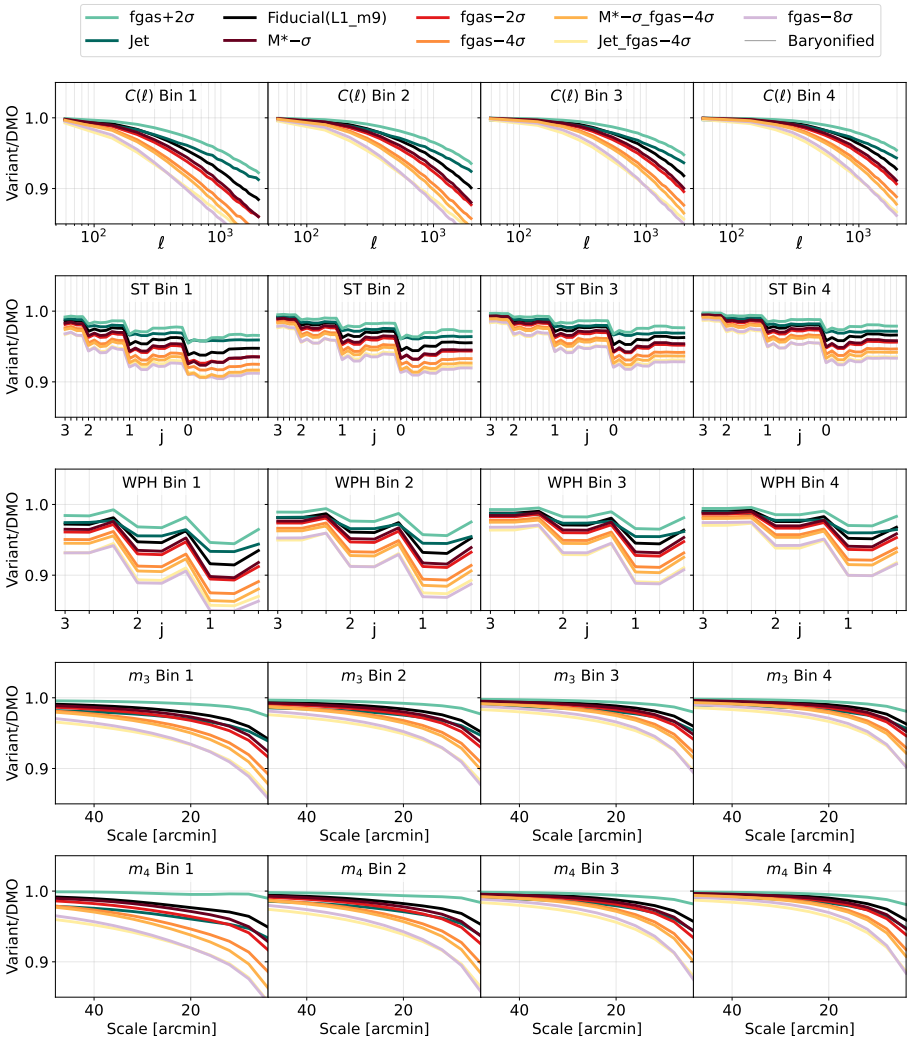

Figure 5.2: Summary statistics of the FLAMINGO κ maps (§ 5.4). Rows correspond to different statistics, and columns to redshift bins (modeled after DES Y3 data). The y-axis shows the FLAMINGO statistics relative to the statistics of the DMO simulation of the same initial condition. On the x-axis, spatial scales decreases to the right. ST and WPH are ordered by the wavelet size (j) as a proxy for their characteristic scale. In FLAMINGO, baryon feedback physics suppresses all statistics on all scales, and is the strongest in low redshift bins.

5.5.2 Scattering Transform Coefficients (ST)

Wavelets are localized oscillatory functions that act as band-pass filters in both real and Fourier space. Unlike Fourier analysis which is a global transformation, wavelets provide spatial localization while simultaneously capturing frequency information. A wavelet family consists of wavelets of different scales and orientations, and can be used to extract multi-scale patterns in images while preserving directional information [269].

Given a template wavelet $\psi(\mathbf{r})$,[8] we can generate new wavelets of different sizes and orientations via

$$\psi_{j,l}(\mathbf{r}) \propto \psi\Big(\frac{1}{2^j} R^{-1}_{l\pi/L}\mathbf{r}\Big)\,, \quad j \in \{\mathbb{Z} \mid 0 \leq j < J\}\,, \quad l \in \{\mathbb{Z} \mid 0 \leq l < L\}\,. \tag{5.3}$$

Here j and l label the scale and the orientation of the derived wavelets. Each wavelet covers $\sim 2^j$ image pixels, and larger j corresponds to larger scales. The operator R^{-1} in the argument is the rotation matrix with the angle of rotation given by the subscript, with the wavelets covering L directions uniformly distributed over $[0, \pi)$.

The scattering transformed κ maps are constructed by iteratively applying the wavelet convolution and the modulus operator [244, 269]. The first and second-order scattering transformed κ maps are

$$\kappa^{\mathrm{ST},(1)}_{j,l} = \big|\kappa \star \psi_{j,l}\big|\,, \tag{5.4}$$

$$\kappa^{\mathrm{ST},(2)}_{j,l,j',l_\Delta} = \Big|\,\big|\kappa \star \psi_{j,l}\big| \star \psi_{j',l-l_\Delta}\Big|\,, \tag{5.5}$$

8. For scattering transform, we use the Morlet wavelet under the convention of [244].

where $\star$ denotes convolution. The modulus preserves the first-order field statistics while introducing the necessary non-linearity to extract information beyond the mean without the instability to outliers that higher-order moments would cause. The higher-order convolution extracts information about spatial distributions $\kappa^{\mathrm{ST},(1)}$, and captures HOS by measuring the "clustering of clustered structures," with each order n potentially containing information up to 2^n-point functions [244, 270]. For the second-order field, we typically require $j \leq j'$ to ensure that the second wavelet extracts information at scales equal to or larger than the first. Since κ is a homogeneous and isotropic field, we compress κ^{ST} through spatial and angular averaging [244, 271]. More specifically, the ST coefficients are given by

$$s_j^{(1)} = \langle \kappa_{j,l}^{\mathrm{ST},(1)} \rangle_{p,l} \,, \tag{5.6}$$

$$s_{j,j',l_\Delta}^{(2)} = \langle \kappa_{j,l,j',l_\Delta}^{\mathrm{ST},(2)} \rangle_{p,l}, \tag{5.7}$$

where $\langle \cdot \rangle$ denotes the expectation value, and the subscripts indicate averaging over pixels (p) and orientations (l). These coefficients compactly and robustly summarize the information in the hierarchical scales of the clustering signal beyond the power spectrum.

To measure ST, we downgrade the κ maps to $\texttt{NSIDE} = 1024$, and divide them into $\texttt{NSIDE} = 8$ HEALPix pixels, each gnomonically projected to a 256×256 Cartesian patch with $3.44'$ resolution. Each patch is centered on an $\texttt{NSIDE} = 8$ pixel with surrounding HEALPix pixels masked to prevent double counting. We measure ST

on 1/8 of the sky using the `scattering_transform` package[9] with parameters $J = 4$ and $L = 3$, leading to 4 $s^{(1)}$'s and 30 $s^{(2)}$'s per redshift bin. The final ST is calculated independently for each tomographic bin and averaged across all Cartesian patches.

Since ST does not have a sharp scale cutoff like $C(\ell)$, we empirically determine the characteristic scale of each ST, and perform consistent scale cut in App. 5.9.1. All 34 ST coefficients per redshift bin are found to predominantly probe scales below $\ell_{\rm max} = 2000$.

5.5.3 Wavelet Phase Harmonics Coefficients (WPH)

WPH is another class of wavelet-based summary statistics. Again, we use $\psi_{j,l}(\mathbf{r})$ to denote wavelets of different sizes (j) and orientations (l), but with a different waveform – the bump-steerable wavelets [272]. WPH statistics are obtained by first convolving κ with $\psi_{j,l}(\mathbf{r})$, then taking its phase harmonics, and finally computing the moments of the resulting fields [245, 272]. Since κ is isotropic, we also average over the orientations of the wavelets. The equations for the first- and second-order WPH reduce to

$$S_j^{(1,1)} = \langle \mathrm{Cov}\big(\kappa \star \psi_{j,l},\, \kappa \star \psi_{j,l}\big)\rangle_l \,, \tag{5.8}$$

$$S_j^{(0,0)} = \langle \mathrm{Cov}\big(|\kappa \star \psi_{j,l}|,\, |\kappa \star \psi_{j,l}|\big)\rangle_l \,, \tag{5.9}$$

$$S_j^{(0,1)} = \langle \mathrm{Cov}\big(|\kappa \star \psi_{j,l}|,\, \kappa \star \psi_{j,l}\big)\rangle_l \,. \tag{5.10}$$

9. `github.com/SihaoCheng/scattering_transform`

$S^{(0,0)}$ and $S^{(0,1)}$ contain both Gaussian and non-Gaussian information due the non-linearity of the phase harmonics (in this case, the modulus) operation. $S^{(1,1)}$ contains only Gaussian information, albeit with a different smoothing scheme compared to the power spectrum.

The measurement of WPH is similar to that of ST, on gnomonically projected patches of $3.44'$ resolution. We again work with Cartesian patches that are projected from the κ maps. On each patch, we compute WPH using the `PyWPH` package[10]. We consider wavelets with $J = 4$ and $L = 3$, leading to 4 of each $S^{(1,1)}$, $S^{(0,0)}$, and $S^{(0,1)}$ per redshift bin. We then average the WPH from each patch across 1/8 of the sky to reduce cosmic variance.

The scale cut for WPH is empirically determined in App. 5.9.1 and is consistent with the $\ell_{\rm max} = 2000$ scale cut for the power spectra. The three $j = 0$ WPH are removed from subsequent analyses.

5.5.4 Higher Order Moments

The third and fourth moments of the convergence field are defined as

$$m_3(\theta) = \left\langle \big(\kappa \star G(\theta)\big)^3 \right\rangle_p , \tag{5.11}$$

$$m_4(\theta) = \left\langle \big(\kappa \star G(\theta)\big)^4 \right\rangle_p , \tag{5.12}$$

where $G(\theta)$ is a Gaussian kernel of width θ and p denotes pixels. m_3 measures skewness of κ. It contains only non-Gaussian information and is strictly zero for a Gaus-

10. `github.com/bregaldo/pywph`

sian random field. m_4 measures the kurtosis of κ and has both a Gaussian (derivable from the power spectrum) and a non-Gaussian contribution following Wick's theorem. We set the minimum scale at $4'$ and use $\theta \in [4', 48']$ in increments of $4'$. This is consistent with the scale cuts of other statistics since a $4'$ Gaussian-smoothed κ has negligible power beyond $\ell = 2000$.

The modeling and the real data measurement of m_3 and m_4 are more sensitive to noise than two-point statistics. This is because 1) by the central limit theorem, for a sample with size N, while $\langle\kappa\rangle$ has an error $\propto N^{-1/2}$, for $\langle\kappa^n\rangle$, the effective standard error scales more like $\alpha N^{-1/2}$ where α grows superlinearly with n, and 2) the tail of outliers of the distribution contributes disproportionally, amplifying the fluctuations. Perturbation theory provides a similar perspective, where empirically $\langle\kappa^n\rangle \propto \langle\kappa^2\rangle^{n-1}$ [273]. Therefore an x uncertainty in $\langle\kappa^2\rangle$ propagates to $(n-1)x$ in $\langle\kappa^n\rangle$.

5.5.5 The statistics of FLAMINGO and baryonified convergence maps

In total, we consider 87 (348) data points per redshift bin (and in total), as summarized in Tab. 5.2. Throughout this work, we consider summaries of the auto-correlation of the four tomographic bins, but do not include their cross-correlations for the sake of reducing the size of the data vector. We still test the ability of the baryonification model to fit all four tomographic bins, and are therefore adequately probing the model's redshift evolution (as it pertains to weak lensing analyses). The statistics for $\kappa_{\textbf{flamingo}}$, relative to those of $\kappa_{\textbf{dmo}}$, are shown in Fig. 5.2. In all

FLAMINGO feedback variants, baryons suppress all statistics at all scales. The suppression is stronger at smaller scales and lower redshifts, consistent with previous works [e.g., 204, 211, 214, 228, 274, 275]. Relative to the fiducial model, the fgas+2σ variant shows weaker baryonic signatures, whereas the fgas−2σ and Jet_fgas−4σ variants induce stronger suppression in the statistics.

Statistics		**Equation**	**N per bin**	**Notes**
ST	$s_j^{(1)}$	Eq. 5.6	4	$0 \leq j < 4$
	$s_{j,j',l_\Delta}^{(2)}$	Eq. 5.7	30	$0 \leq j \leq j' < 4$, $0 \leq l_\Delta < 3$
WPH	$S^{(1,1)}$	Eq. 5.8	3	$1 \leq j < 4$
	$S^{(0,0)}$	Eq. 5.9	3	$1 \leq j < 4$
	$S^{(0,1)}$	Eq. 5.10	3	$1 \leq j < 4$
$C(\ell)$		Eq. 5.2	20	linear bin, $\ell \in [20, 2020]$
$m_3(\theta)$		Eq. 5.11	12	smoothing scale $\theta \in [4', 48']$
$m_4(\theta)$		Eq. 5.12	12	smoothing scale $\theta \in [4', 48']$

Table 5.2: Summary of statistics used in the analysis, restricted to scales with $\ell \lesssim 2000$.

To build intuition about how each BCM parameter affects the statistics, we vary each parameter independently over its prior range defined in Tab. 5.1, and plot the resulting statistics of $\kappa_{\mathbf{bary}}$ relative to those of $\kappa_{\mathbf{dmo}}$ in Fig. 5.3. Our final results in § 5.7 suggest that the three most important parameters to match and capture variations among different FLAMINGO feedback models are M_c, $\theta_{\rm ej}$, and η, which we highlight in Fig. 5.3. The effects of the remaining parameters are shown in Fig. 5.8 in App. 5.9.2. We find that the prior ranges induce a wide range of changes in the statistics—including both suppression and, in some cases, enhancement—relative to

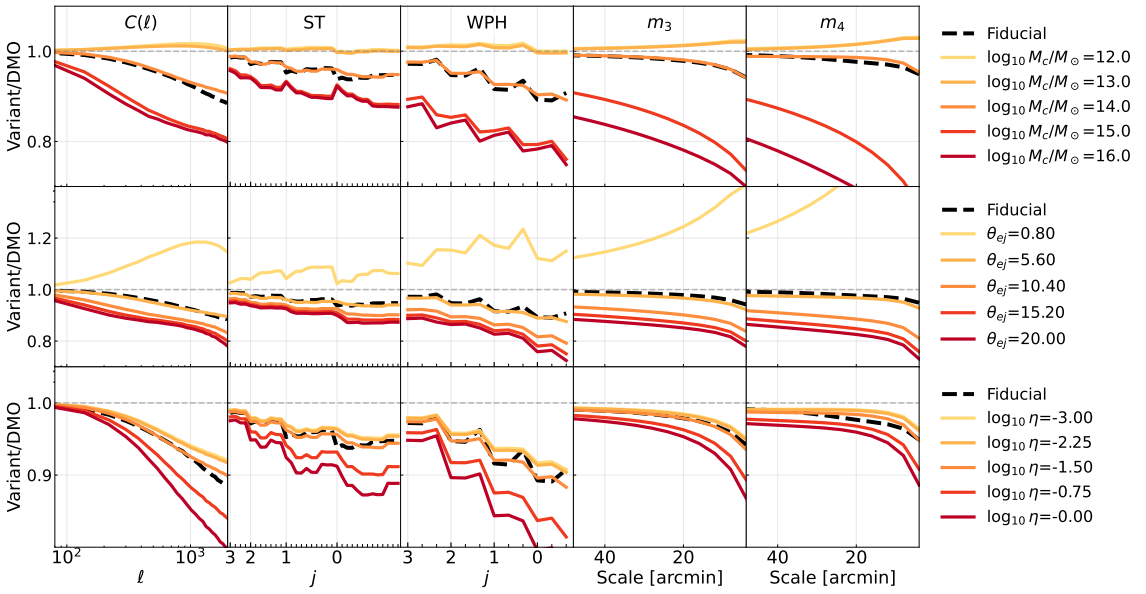

Figure 5.3: Summary statistics of the baryonified (colored) and the FLAMINGO fiducial variant's (black, dashed) convergence map, focusing on the first redshift bin. For the baryonified maps, a single BCM parameter is varied at a time. Rows correspond to different BCM parameters, and columns to different statistics. The y-axis shows the statistics of both the baryonified and FLAMINGO maps, relative to those of the DMO simulation with the same initial condition. On the x-axis, spatial scale decreases to the right. ST and WPH statistics are ordered by wavelet size (j), used here as a proxy for characteristic scale. The BCM flexibly generates a range of baryonic signatures in the statistics, clustering around the FLAMINGO results. The variation of other parameters in Tab. 5.1 is shown in Fig. 5.8 in App. 5.9.2.

the fiducial variant' s measurement.

5.6 Method

We want to demonstrate that the BCM can flexibly and accurately reproduce the wide range of baryonic feedback models in FLAMINGO, in terms of two-point and HOS and investigate what are the baryon parameter values that best fit the

FLAMINGO simulations and how they vary as a function of hydrodynamical feedback strengths.

In principle, we could sample the best-fit baryon parameters in one step using a Markov Chain Monte Carlo (MCMC) – given the baryon parameters, we could generate baryonified κ maps using the DMO simulation, measure their statistics, and compare them to those from FLAMINGO. In practice, generating new baryonified maps and computing their statistics at each MCMC step is too time-consuming. Therefore, we proceed in three steps:

1. Build an emulator that predicts the statistics for the given baryon parameters
2. Use the emulator in an MCMC on each of the FLAMINGO simulations to find the best-fit baryon parameters
3. Once the MCMC chains have converged, we generate baryonified κ maps at the MCMC best fit, measure the suppression of the statistics, and use them for our final considerations.

Therefore, we use the emulator as a fast and efficient tool to refine our search in parameter space, rather than brute-forcing the problem by baryonifying the maps within the MCMC, which would be computationally infeasible. This means that possible discrepancies between the suppression of the statistics in the FLAMINGO simulations and the one we obtain at the end of our three-step procedure can be attributed to two factors: (1) the accuracy of the emulator in reproducing the statistics compared to the full baryonification process, and (2) the intrinsic limitations of the baryonification method itself, even assuming a perfect emulator.

5.6.1 Emulator

We will refer to the collection of all summary statistics as $\mathcal{S}$. To accelerate the MCMC inference, we construct an emulator r, which predicts the ratio of the BCM-corrected summary statistics to its DMO counterpart given baryon parameters $\mathbf{b}$, i.e.,

$$r(\mathbf{b}) = \mathcal{S}_{\mathbf{bary}}(\mathbf{b})/\mathcal{S}_{\mathbf{dmo}}, . \tag{5.13}$$

The numerator $\mathcal{S}_{\mathbf{bary}}(\mathbf{b})$ is computed by first baryonifying $\kappa_{\mathbf{dmo}}$ using the BCM (§ 5.3) and then performing two-point and HOS measurements (§ 5.5). Since $\kappa_{\mathbf{bary}}$ and $\kappa_{\mathbf{dmo}}$ share the same initial condition, the ratio eliminates shot noise and cosmic variance.

The training set includes $5,468$ $\kappa_{\mathbf{bary}}$ maps sampled in the ten-dimensional baryon parameter space. The prior ranges for the parameters are detailed in Tab. 5.1 and are similar to those used in Anbajagane *et al.* [3], Schneider *et al.* [234]. In Figs. 5.3 and 5.8, we have shown how varying a single BCM parameter can produce suppression similar to those observed in the fiducial model. We also explicitly verify that, across the training samples, the baryonified maps exhibit baryonic signatures spanning the full range of variations seen in the different FLAMINGO feedback variants. For each statistic, we construct an ensemble of 10 neural networks via cross-validation. We use the mean and 1σ scatter of the ensemble predictions as the point estimate and uncertainty of $r(\mathbf{b})$. The emulator accuracy is shown in Fig. 5.4 as a function of scale and redshift, and generally decreases at smaller scales and lower redshifts, where baryonic feedback is more nonlinear and varied. The emulator achieves a mean

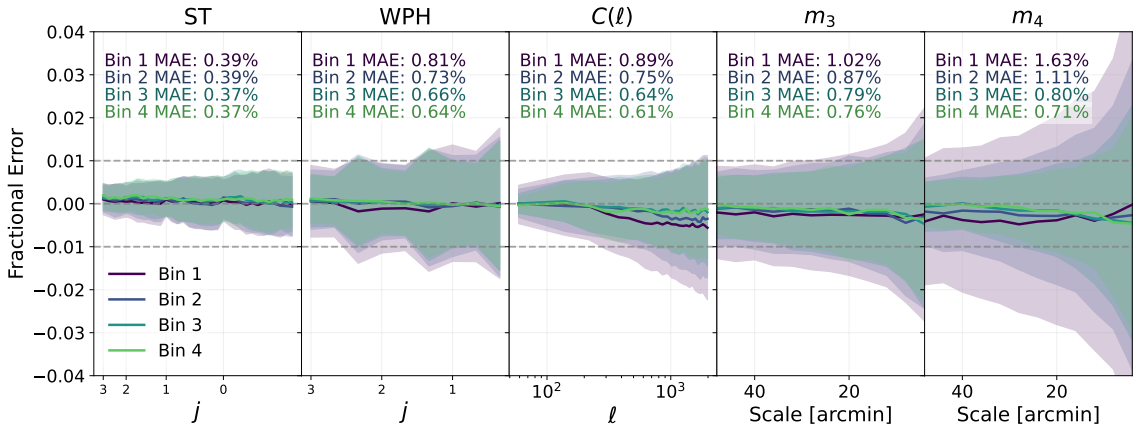

Figure 5.4: The emulator mean relative error, grouped by statistics and by redshift bins. The mean absolute (relative) error (MAE) is shown in text. The error is evaluated by averaging across the test set across the baryon parameter space. The emulator achieves sub-percent accuracy for all statistics besides the low redshift moments. The higher-order moment statistics are intrinsically nosier since it involves multiplications of random variables. We remind the reader that the emulator is used to efficiently explore the BCM parameter space when comparing to FLAMINGO statistics. However, the final assessment of how well the BCM reproduces a given FLAMINGO statistic is based on running the full baryonification pipeline at the emulator-derived best-fit parameters.

absolute error below 1% for all statistics except the low-redshift moments, which are intrinsically noisier, as discussed in § 5.5.4. The emulator is implemented in `jax` and supports automatic differentiation [276]. Details of the training set and emulator are provided in App. 5.9.3.

5.6.2 Sampling

For each FLAMINGO feedback variant with statistics $\mathcal{S}_{\mathbf{hydro}}$, we seek the best-fit baryon parameters by sampling the posterior $\mathcal{P}$

$$\mathcal{P}(\mathbf{b} \mid \mathcal{S}_{\mathbf{hydro}}) = \mathcal{G}\left(\frac{\mathcal{S}_{\mathbf{hydro}}}{\mathcal{S}_{\mathbf{dmo}}} - r(\mathbf{b}), \sigma^2 + \epsilon(\mathbf{b})^2\right) \tag{5.14}$$

where $\mathcal{G}$ is a Gaussian distribution parametrized by mean and variance, σ is a constant uncertainty per statistic and ϵ is the emulator uncertainty. We assume a flat prior on $\mathbf{b}$. As discussed above, the simulations are noiseless and the ratio statistics suppress the cosmic variance and shot noise. Therefore, in theory, the emulator uncertainty ϵ is the only source of error in $r(\mathbf{b})$. However, since ϵ is a function of $\mathbf{b}$, the parameter-dependent covariance makes the sampling very difficult in high-dimensional space. Therefore, we introduce a constant noise σ to promote the smoothness of the posterior surface to facilitate sampling. In particular, σ does not reflect real observational uncertainty. We find $\sigma = 0.04$ for ST, WPH, and $C(\ell)$ and $\sigma = 0.08$ for m_3 and m_4 make the chains converge robustly. Although we are putting less weight on the moments, we shall see that the final best-fit baryon parameters reproduce all statistics equally well. We leverage the gradient information provided by the emulator and use the `numpyro` implementation [58, 59] of the Hamiltonian Monte Carlo algorithm for efficient sampling [17, 27].

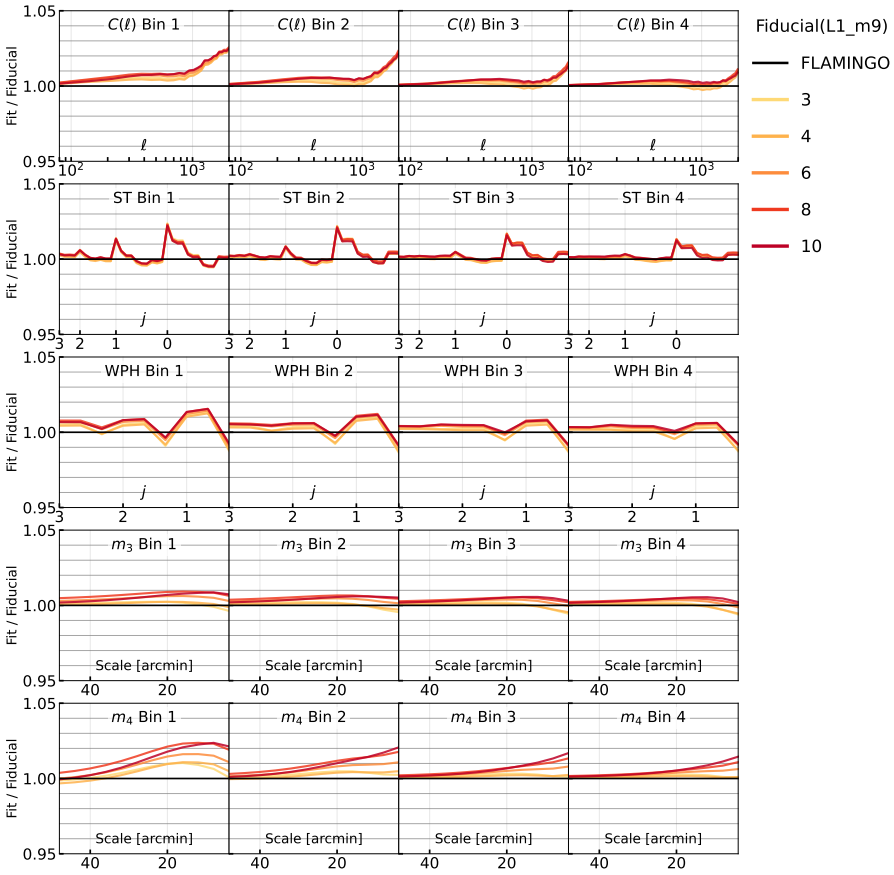

Figure 5.5: The residuals error of the BCM when fitted to the FLAMINGO fiducial variant. Rows correspond to different statistics, and columns to redshift bins. The y-axis shows the ratio between the statistics of the best-fit baryonified map to that of the fiducial variant. On the x-axis, spatial scales decreases to the right. ST and WPH are ordered by the wavelet size (j) as a proxy for their characteristic scale. Colors indicate the number of free parameters in the BCM fit. The shaded region marks $\pm 2\%$ for visual reference. For all n-parameter models, the BCM flexibly reproduces the baryonic features in the FLAMINGO fiducial variant simulation to percent level. The prediction is robust even with only 3 degrees of freedoms. Differences in accuracy between the models stem from how we estimate the best-fit parameter, as discussed in § 5.7.2. The residuals for the strongest (fgas-8σ) and weakest (fgas$+2\sigma$) feedback variants are shown in Fig. 5.9 and Fig. 5.10 in App. 5.9.3.

5.7 Results

5.7.1 The 10-parameter baryon correction model

We sample the posterior $\mathcal{P}(\mathbf{b} \mid \mathcal{S}_{\mathbf{hydro}})$ across the full 10-dimensional baryon parameter space for each of the FLAMINGO feedback variants. Since the marginal posterior distributions are approximately Gaussian, we use the medians – rather than the maximum a posteriori (MAP) values, which are less robust against sample variance – as our best-fit parameters $\hat{\mathbf{b}}$. The estimated $\hat{\mathbf{b}}$ values and their 1σ uncertainties are listed in Tab. 5.4 in App. 5.9.4.

We then investigate how well the best-fit BCM reproduces the statistics in $\kappa_{\mathbf{flamingo}}$. Ideally, we would directly compare the emulator prediction $r(\hat{\mathbf{b}})$ with FLAMINGO. However, the emulator includes bias from both the neural network and the BCM itself. To isolate the latter, we directly construct $\kappa_{\mathbf{bary}}(\hat{\mathbf{b}})$, measure its statistics, and compare them to the FLAMINGO results. The emulator results agree with direct baryonification within expected emulator errors. For example, for FLAMINGO's fiducial variant, the relative error of the 10-parameter best-fit BCM is shown in Fig. 5.5 in dark red. The BCM accurately reproduces ST, WPH, $C(\ell)$, and moment statistics at all scales and redshifts to within 2%. The residual error is the largest at small scales and low redshifts, where baryonic feedback is strongest.

We analyze all 8 other FLAMINGO variants in the same way and show the residuals for the strongest (fgas-8σ) and weakest (fgas$+2\sigma$) FLAMINGO feedback variants in Fig. 5.9 and Fig. 5.10 in App. 5.9.3. The BCM can reproduce the statistics of different FLAMINGO variants equally well. For the strongest AGN feedback

case (fgas-8σ), the BCM shows elevated errors of 3–4% in the first redshift bin of m_4 for $\theta < 20$ arcminutes, but otherwise maintains 2% accuracy. However, as discussed in § 5.5.4, higher-order moments are intrinsically noisier in both modeling and measurement. The observed m_4 discrepancy remains well below the statistical uncertainty of a DES-like survey [15] and is only detected at low significance in current DES Y3 data [215].

We have verified that a significant fraction of the remaining 2% residual at small scales can be alleviated by lowering the halo mass threshold to $10^{12.5}\,M_\odot/h$. This test was performed using the best-fit of the baryonification emulator on the Fiducial simulation, and should be considered an approximate check. A formal evaluation would require regenerating the maps, retraining the emulator, and refitting the model using $10^{12.5}\,M_\odot/h$ as the threshold —a process we do not pursue due to its substantial computational cost. Nonetheless, we note that adopting a threshold of $10^{13}\,M_\odot/h$ still yields sub-2% precision.

Thus far, we have investigated the BCM at the fiducial cosmology. While both $\mathcal{S}_{\mathbf{hydro}}$ and $\mathcal{S}_{\mathbf{dmo}}$ depend on cosmology and astrophysics, these dependencies are approximately factorizable [204, 277]. As a result, the cosmology dependence largely cancels in the ratio $\mathcal{S}_{\mathbf{hydro}}/\mathcal{S}_{\mathbf{dmo}}$, which the BCM models. Using the low-S_8 variants (LS8, LS8_fgas-8σ) of the FLAMINGO suite, we show in App. 5.9.3 that this ratio varies by less than 1% compared to the fiducial cosmology even for the strong feedback model (fgas-8σ). Since the BCM (and thus the emulator) predicts the suppression ratio solely from the baryonic parameters, the recovered LS8 BCM parameters are consistent with those from the fiducial cosmology.

The baryonic suppression has a stronger dependence on the baryon fraction, Ω_b/Ω_m [e.g. 234] but we leave a detailed analysis of this dependence to future work due to the computational cost of training separate emulators for each cosmology.

To summarize, for $\ell \leq 2000$, the 10-parameter BCM is flexible enough to reproduce the range of baryonic signatures in two-point and HOS.

5.7.2 Baryon correction model with restricted degrees of freedoms

Variant	$\log_{10} M_c/M_\odot$	$\log_{10} \eta$	$\theta_{\rm ej}$
Fiducial(L1_m9)	13.9 ± 0.19	-1.51 ± 0.39	4.80 ± 0.65
Jet	13.7 ± 0.16	-1.87 ± 0.39	$6.23^{+0.89}_{-0.75}$
Jet_fgas-4σ	14.0 ± 0.16	$-1.02^{+0.21}_{-0.25}$	5.94 ± 0.61
fgas$+2\sigma$	13.7 ± 0.16	-1.95 ± 0.38	$5.05^{+0.68}_{-0.59}$
fgas-2σ	14.0 ± 0.22	$-1.20^{+0.34}_{-0.35}$	4.85 ± 0.59
fgas-4σ	14.0 ± 0.19	$-0.983^{+0.23}_{-0.29}$	5.05 ± 0.54
fgas-8σ	14.0 ± 0.14	$-0.885^{+0.17}_{-0.18}$	5.40 ± 0.52
M*$-\sigma$	14.0 ± 0.21	$-1.17^{+0.33}_{-0.37}$	4.69 ± 0.55
M*$-\sigma$_fgas-4σ	14.0 ± 0.15	-0.852 ± 0.2	5.05 ± 0.5

Table 5.3: The best-fit BCM parameters obtained from the joint fitting of all HOS to different FLAMINGO variants. These parameter combinations yield HOS that agree with FLAMINGO measurements to 2%. Other BCM parameters are fixed to $\nu_{M_c} = 0$, $\nu_{\theta_{\rm ej}} = 0$, $\gamma = 2.60$, $\log_{10}\theta_{co} = -1.00$, $\log_{10}\eta_\delta = -0.28$, $\delta = 14.0$, and $\mu_\beta = 1.60$. The uncertainty is the 16% and 84% credible interval of our fitting method. See Tab. 5.4 in App. 5.9.4 for the best-fit values for all n-parameter models.

After validating that the 10-parameter BCM accurately reproduces the statistics, we ask whether similar success can be achieved with significantly fewer degrees of freedom. This simplification has three advantages. First, in real survey analyses, using fewer baryon parameters reduce the impact on the constraining power

of key cosmological parameters. Second, in simulation-based inference, it lowers the number of simulations required to train the neural posterior estimator, thus reducing computational cost. Third, the 10-dimensional $\mathbf{b}$ space contains significant degeneracies; fewer parameters improves the interpretability how $\mathbf{b}$ vary across FLAMINGO feedback variants.

We consider 4 restricted versions of the BCM, with 8, 6, 4, and 3 parameters. In general, at each step, we fix parameters whose best-fit values are consistent across all feedback variants. The 8-parameter model fixes the redshift scalings ν_{M_c} and $\nu_{\theta_{\rm ej}}$ to zero, consistent with their posteriors in the 10-parameter fits. The 6-parameter model further fixes $\gamma = 2.6$ and $\log_{10}\theta_{co} = -1$. For the 4-parameter model, we also fix $\log_\delta \eta_\delta = 0.28$ and $\delta = 14$. Finally, for the 3-parameter model, we additionally fix $\mu_\beta = 1.6$. The remaining free parameters in the 3-parameter model are $\log_{10} M_c$, $\log_{10}\eta$, and $\theta_{\rm ej}$. For each n-parameter model and feedback variant, we tabulate $\hat{\mathbf{b}}$ in Tab. 5.4 in App. 5.9.4, with fixed parameters shown in red. We also summarize $\hat{\mathbf{b}}$ values for the 3-parameter model in Tab. 5.3. As with the 10-parameter case, we validate the best-fit BCMs by directly constructing $\kappa_{\mathbf{bary}}$ and measuring their statistics. Residuals for the fiducial, strong feedback (fgas-8σ), and weak feedback (fgas$+2\sigma$) variants are shown in Figs. 5.5, 5.9, and 5.10, respectively.

Although we have dramatically reduced the number of parameters, all n-parameter models faithfully reproduce all statistics across scales, redshifts, and feedback variants. Notably, the 3-parameter model often fits the statistics better than the 10-parameter model. This is because our best estimate, $\hat{\mathbf{b}}$, is defined as the median of the marginal posterior. In high-dimensional spaces, $\hat{\mathbf{b}}$ can deviate from the MAP due

to projection over non-Gaussian volumes. As the number of parameters decreases, the posterior becomes more Gaussian-like, and $\hat{\mathbf{b}}$ approaches the MAP, leading to improved fits. In other words, were we to have a less noisy emulator, we could potentially use the MAP as $\hat{\mathbf{b}}$ and obtain even higher-quality fits for the 10-parameter model.

5.7.3 Baryon parameters

Having demonstrated that the BCM can flexibly reproduce the baryonic signatures across a range of FLAMINGO feedback variants, we now examine the best-fit baryon parameters themselves. When comparing with other results in the literature, we emphasize that our modeling choices —such as the specific implementation of the BCM and the set of free parameters —are in some aspects different from those adopted in other studies, and that our results are entirely based on simulations. In this context, consistency at the parameter level with earlier work is not required for our approach to be effective, as success is defined by how well it reproduces the suppression of the lensing statistics in the FLAMINGO simulations, rather than by agreement in individual parameter values. Nonetheless, the inferred $\hat{\mathbf{b}}$ is broadly consistent with earlier results based on both simulations and observations [234, 236, 239]. This qualitative agreement reinforces that our model is not overfitting to a narrow region of parameter space and is instead capturing physically meaningful baryonic effects.

Focusing on the 10-parameter model, all variants are consistent with $\nu_{M_c} = 0$ and $\nu_{\theta_{\rm ej}} = 0$, suggesting minimal sensitivity of our measurements to the mass dependence

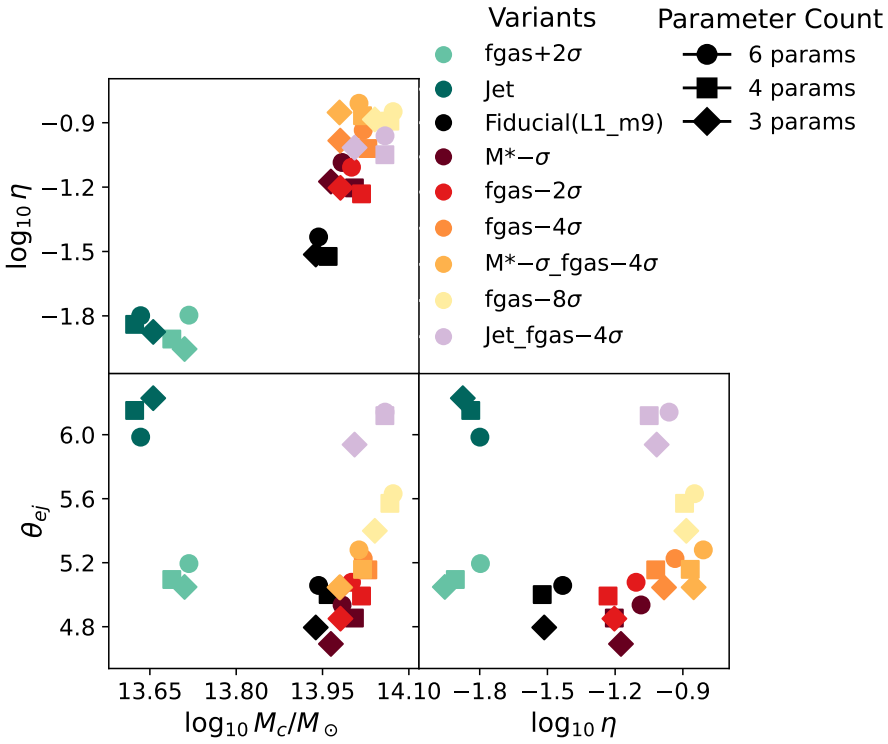

Figure 5.6: The inferred baryon parameter $\hat{\mathbf{b}}$ from different FLAMINGO feedback variants. Markers represent inference using different degrees of freedom. The color map labels the feedback variants and inherits from Fig. 5.2. In the legend, the feedback strength approximately increases from top to bottom, with the strongest (weakest) feedback model suppression being the fgas-8σ (fgas$+2\sigma$) model. For each feedback variant, the clustering of the points demonstrate the BCM is robust against removing many nuisance parameter dependencies.

of the gas ejection radius and characteristic slope. This lack of sensitivity is consistent with the forecasted constraints from weak lensing presented in [3]. Interestingly, all feedback variants prefer a high δ, which corresponds to a sharp drop off in the gas density profile past $r > \theta_{\rm ej} R_{\rm 200c}$ [3, 234, 239]. To investigate if different values of δ affect the best-fit statistics, we rerun the 4- and 3-parameter models, setting $\delta = 7$ and $\mu_\beta = 2$ while keeping the other fixed parameters the same as in § 5.7.2. The BCM still achieves a 2% fit across all statistics and feedback variants, although the best-fit $\theta_{\rm ej}$ lowers from 5 to 3. This is expected as δ and $\theta_{\rm ej}$ both control the radial extent of the gas profile and are therefore quite negatively correlated.

Next, we investigate how much the estimated $\hat{\mathbf{b}}$ varies across the n-parameter models. Fig. 5.6 shows $\hat{\mathbf{b}}$ as a function of feedback variant for the 6-, 4-, and 3-parameter models. The constraints on the remaining free parameters do not vary significantly as we reduce the number of parameters. This demonstrates that the BCM's weak lensing predictions are robust against removing many nuisance parameters and the BCM allows us to track how the phenomenological baryon parameters change as a function of FLAMINGO feedback models. The best-fit parameters of the different models show some degree of clustering in the 3D parameter space, indicating that, if sufficient statistical power is achieved by observational probes, it may be possible to distinguish between feedback models using real data. Concerning the physical interpretation of the position of the models in the parameter space, we know that $\log_{10} M_c$, $\log_{10} \eta$, and $\theta_{\rm ej}$ all scale positively with feedback strength. Indeed, we find that the strongest feedback model (fgas-8σ) lies in the upper range of each of the three parameters, while the weakest feedback model (fgas$+2\sigma$ weak) occupies

the lowest values in each dimension. Beyond a qualitative ranking of suppression strength, the interplay between the three parameters and the resulting impact on the lensing statistics is non-trivial. Each parameter modulates the suppression differently and in a scale-dependent way, leading to best-fit values that are scattered across parameter space in a complex pattern.

5.8 Conclusion

Data from current and upcoming weak lensing surveys are probing cosmological structures at increasingly small spatial scales. To extract both cosmological and astrophysical information in this regime robustly and optimally, we require models that can accurately capture the full range of baryonic feedback effects allowed by theory and observed in state-of-the-art hydrodynamical simulations.

In this work, we investigate the map-level baryonification approach from Anbajagane *et al.* [3], which efficiently and flexibly models baryonic effects in weak lensing convergence maps. By calibrating the baryon parameters against FLAMINGO hydrodynamical simulations, we demonstrate that this baryon correction model (BCM) accurately reproduces baryonic impacts on various convergence field statistics, including the two-point power spectrum, scattering transform coefficients (ST), wavelet phase harmonics (WPH), and third- and fourth-order moments. This agreement holds across scales $\ell < 2000$ and throughout the redshift range of DES-like surveys. We construct an emulator to predict BCM-generated statistics efficiently, then sample baryon parameters by comparing emulator predictions with measurements from different FLAMINGO feedback variants. Testing BCM versions with 10, 8, 6, 4, and

3 free parameters, we find all models can reproduce statistics within $< 2\%$ accuracy with accuracy improving at larger scales and higher redshifts. A minor exception is the fourth moment, where, for models with more than 3 free parameters, the residual error approaches 3% around $10'$, partially due to the increased uncertainty in our fitting procedure.

These results have several implications for future weak lensing analyses. First, this shows that the BCM is robust for a wide range of weak lensing statistics and can be embedded in simulation-based inference frameworks to marginalize over baryonic uncertainties or directly constrain baryon parameters. Reduced-parameter versions of the BCM further improve efficiency by lowering the number of simulation needed and sampling costs. Second, our work can be extended to test the robustness of other tracers and probes, such as cross-correlations with large-scale structure observables like galaxy clustering or thermal Sunyaev-Zel' dovich maps. Joint analyses with these additional datasets may offer tighter constraints on baryonic physics.

Although our results are based on the DES Y3 redshift distributions, they are broadly applicable to other current weak lensing surveys due to similar source redshift distributions [1, 278, 279]. They are also relevant for upcoming surveys like LSST, Euclid, and Roman, whose higher-redshift distributions may allow our models to extend to even smaller angular scales due to weaker baryonic feedback at early times. The ability to model baryonic physics in this regime opens new opportunities for extracting both cosmological and astrophysical information from small-scale data. Future work could extend this framework to alternative summary statistics, such as voids, peaks, or CNN-based approaches [e.g., 195, 280]. Our results highlight the

promise of map-level baryonification—without relying on 3D particle snapshots—as a computationally efficient and physically grounded approach for incorporating baryonic feedback in weak lensing analyses. As upcoming surveys push to smaller scales, such methods will be essential for maximizing constraining power while remaining robust to astrophysical systematics.

Acknowledgments

AZ is supported by Jane Street through the Jane Street Graduate Research Fellowship. MG is supported by funds provided by the Kavli Institute for Cosmological Physics at the University of Chicago through an endowment from the Kavli Foundation. DA is supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE 1746045. This work was supported by FermiForward Discovery Group, LLC under Contract No. 89243024CSC000002 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics. This work used the DiRAC@Durham facility managed by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility (`www.dirac.ac.uk`). The equipment was funded by BEIS capital funding via STFC capital grants ST/K00042X/1, ST/P002293/1, ST/R002371/1 and ST/S002502/1, Durham University and STFC operations grant ST/R000832/1. DiRAC is part of the National e-Infrastructure.

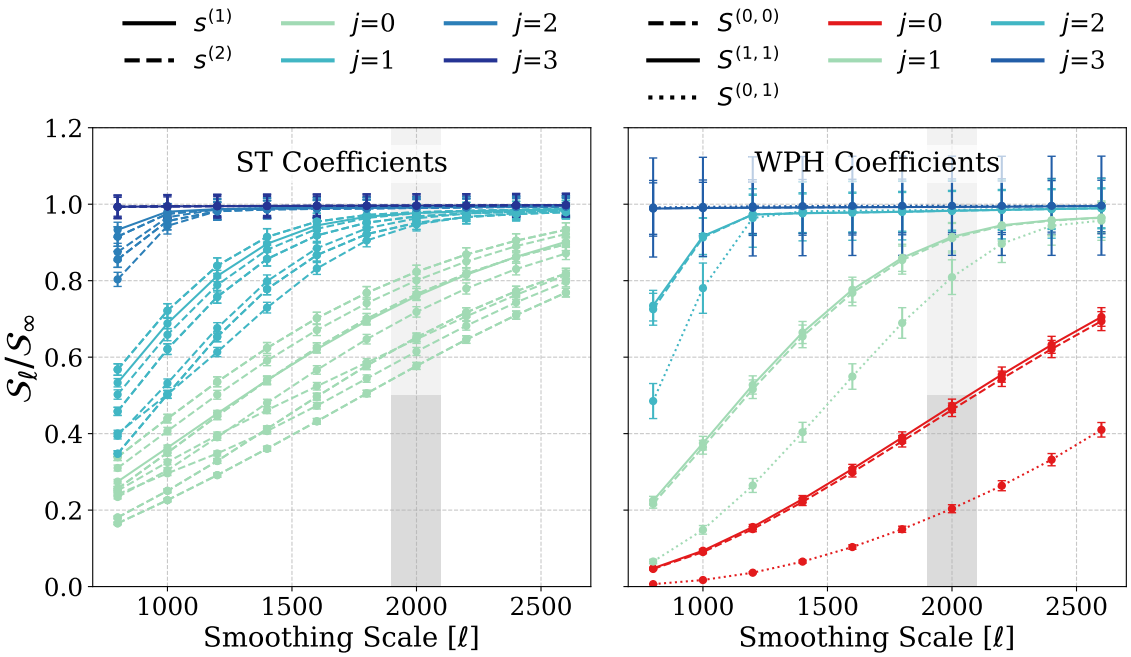

Figure 5.7: The effect of smoothing scale on the two kinds of wavelet coefficients, ST and WPH. Given a statistics $\mathcal{S}$, we define $\mathcal{S}_\ell$ as the statistics measured on κ maps that are low-pass filtered (with a step function) at different ℓ. The statistics measured on the smoothed map is called $\mathcal{S}_\infty$. We test this idea on $\kappa_{\mathbf{dmo}}$, and plot the relative suppression $\mathcal{S}_\ell/\mathcal{S}_\infty$ as a function of smoothing scale ℓ. The line and the error bar represent the mean and the standard error of $\mathcal{S}$ measured on different patches across the full sky. We further color code the lines using the characteristic scale of the wavelet (higher j equals larger wavelets). If $\mathcal{S}_\ell/\mathcal{S}_\infty \ll 1$, then $\mathcal{S}$ contains information beyond ℓ. The $j = 0$ WPH coefficients (shown in red) are suppressed more than 50% at $\ell = 2000$ (shaded region), and hence removed from subsequent analyses.

5.9 Appendix

5.9.1 Scale cut

We empirically determine the characteristic scale of each ST and WPH statistic and enforce a consistent scale cut with respect to the power spectrum. For each statistic $\mathcal{S}$, we smooth the $\texttt{NSIDE} = 1024$ $\kappa_{\mathbf{dmo}}$ maps with low-pass filters at different $\ell_{\rm max}$ values and measure the statistics ($\mathcal{S}_\ell$), normalized by the unsmoothed map's statistics ($\mathcal{S}_\infty$). The results for ST and WPH are shown in the left and right panels of Fig. 5.7, respectively. For a given statistic, if $\mathcal{S}_\ell/\mathcal{S}_\infty \ll 1$, then it contains information beyond ℓ, and vice versa. We exclude statistics where $\mathcal{S}_\ell/\mathcal{S}_\infty < 0.5$ at $\ell_{\rm max} = 2000$. All 34 ST coefficients per redshift bin remain within our analysis scales. For WPH, the three $j = 0$ coefficients (shown in red) fail the scale cut criterion and are removed from the analysis.

5.9.2 Statistics as a function of baryon parameters

Here we show how the summary statistics vary with the remaining BCM parameters listed in Tab. 5.1, excluding M_c, $\theta_{\rm ej}$, and η, which are highlighted in the main text (Fig. 5.3). Each parameter is varied individually across its prior range, and the resulting baryonified statistics are compared to those from the corresponding DMO simulation. These results show the flexibility of the BCM in reproducing a wide range of baryonic effects around the hydrodynamical predictions.

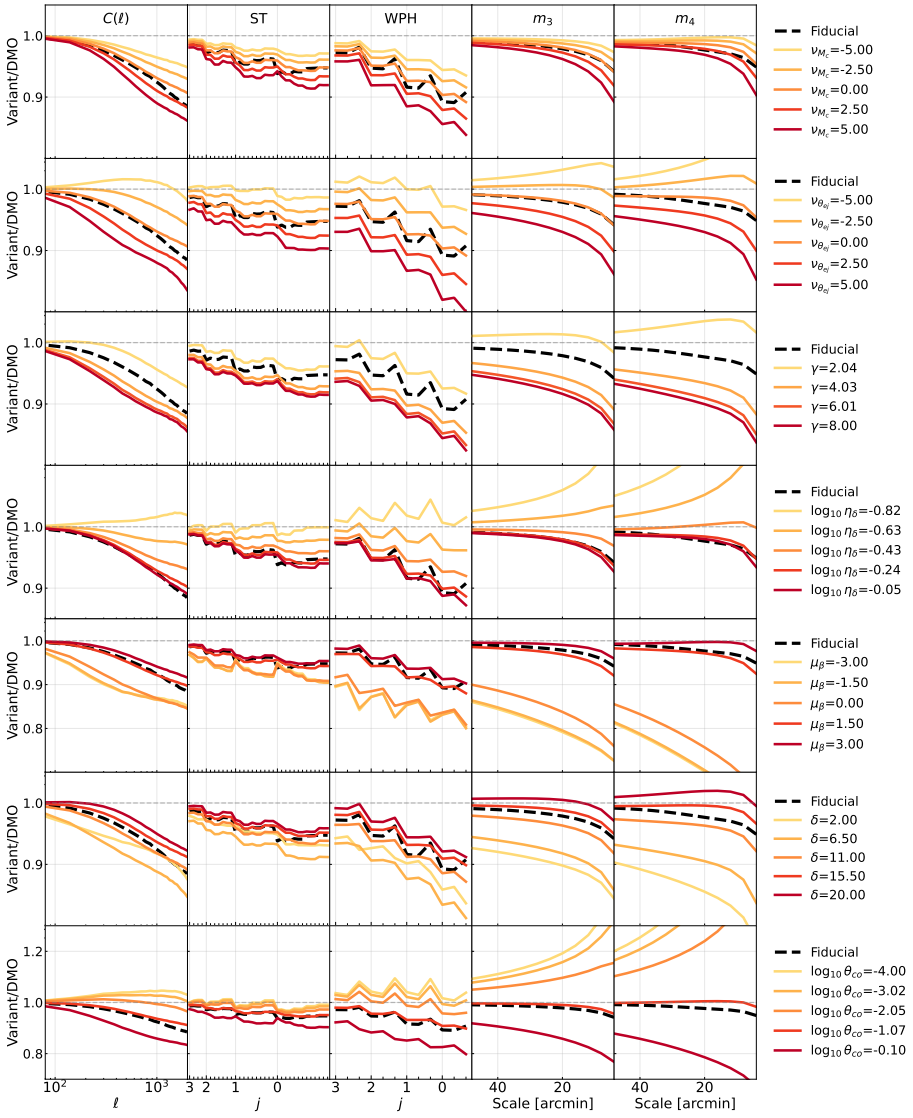

Figure 5.8: Same as Fig. 5.3, but for the remaining BCM parameters not shown in the main text. Colored curves show the summary statistics of baryonified convergence maps with one parameter varied at a time, while black dashed curves show the corresponding statistics from the FLAMINGO simulations.

5.9.3 Emulator

We build a neural network-based emulator to map the baryon parameters to the tomographic ST, WPH, $C(\ell)$ and the m_3 and m_4 statistics. Each type of statistic is modeled independently. We first standardize the input parameters and then use three dense hidden layers following the architectures: (512, 256, 256) for ST, (512, 256, 128) for WPH, (512, 256, 128) for $C(\ell)$ statistics, and (512, 256, 128) for the concatenated m_3 and m_4 data vector. Each hidden layer is followed by a leaky ReLU activation function ($\alpha = 0.1$) [281] and a 3% dropout layer to prevent over-fitting [282]. For robustness, we train 10 independent cross-validated models for each type of statistics, resulting in a total ensemble of 40 neural networks. We use 95% of the data for the 10-fold cross-validation and use the rest as the test set for the performance evaluation as shown in Fig. 5.4. During training, we use the mean absolute error (MAE) as the loss function and the Adam optimizer with weight decay for l_2 regularization [283]. The emulator implements a dual early stopping strategy to prevent over-fitting: (1) a global early stopping mechanism with a patience of 80 iterations that monitors validation loss to terminate training when no improvement is observed, and (2) a learning rate scheduler that starts with a learning rate of 10^{-3} and reduces it by a factor of 0.5 after 5 epochs without improvement, with a minimum threshold of 10^{-6}. Since each statistic is predicted 10 times by the cross-validated models, our emulator not only provides a point estimate (given by the ensemble mean) but also uncertainty estimates (given by the ensemble variance). The entire implementation supports automatic differentiation through `jax`, enabling gradient-based optimization and sampling, as well as sensitivity analysis of the statistics with

respect to the baryon parameters [276].

We used the baryonification model to generate 6695 $\kappa_{\mathbf{bary}}(\mathbf{b})$ maps and measured their statistics $\mathcal{S}_{\mathbf{bary}}$, where the 10-dimensional $\mathbf{b}$ is sampled using a Sobol sequence. Since some $\mathcal{S}_{\mathbf{bary}}$ show an amplification rather than a suppression as observed in all FLAMINGO measurements, we remove $\mathcal{S}_{\mathbf{bary}}$ where $C(\ell)$ is on average amplified by 10% across all linear bins. The final training set includes 5468 $\mathcal{S}_{\mathbf{bary}}$. Since the training set is small, the emulator has high outlier rate near parameter boundaries. Therefore, we truncate parameter domain boundaries where the absolute error averaged across all statistics is above 1%. The most significant truncation occurs for $\gamma < 0.9$. We later verify these restrictions do not impact the posteriors of the MCMC chains.

5.9.4 *Best-fit baryon parameters*

Table 5.4: The best-fit BCM parameters obtained from the joint fitting of all HOS to different FLAMINGO variants. These parameter combinations yield HOS that agree with FLAMINGO measurements to 2%. The integers following the feedback variant names are the number of free parameters. The fixed parameters are highlighted in red. The uncertainty is the 16% and 84% credible interval.

Variant	ν_{M_c}	$\nu_{\theta_{\rm ej}}$	γ	$\log_{10}\theta_{co}$	$\log_{10}\eta_\delta$	δ	μ_β	$\log_{10} M_c/M_\odot$	$\log_{10}\eta$	$\theta_{\rm ej}$
Fiducial(L1_m9)(10)	$1.48^{+2.0}_{-2.4}$	$-0.727^{+1.1}_{-0.99}$	$2.61^{+1.7}_{-0.81}$	-1.06 ± 0.19	$-0.247^{+0.12}_{-0.16}$	13.7 ± 3.5	1.81 ± 0.53	13.9 ± 0.25	-1.57 ± 0.43	$5.36^{+2.4}_{-1.7}$
Fiducial(L1_m9)(8)	0	0	$2.64^{+1.8}_{-0.85}$	-1.04 ± 0.18	$-0.255^{+0.13}_{-0.18}$	$14.4^{+3.1}_{-3.7}$	1.77 ± 0.56	13.9 ± 0.18	$-1.49^{+0.48}_{-0.44}$	$4.85^{+1.8}_{-1.3}$
Fiducial(L1_m9)(6)	0	0	2.60	-1.00	$-0.282^{+0.13}_{-0.17}$	$15.1^{+2.8}_{-3.9}$	$1.87^{+0.56}_{-0.48}$	13.9 ± 0.18	-1.43 ± 0.46	$5.06^{+0.91}_{-1.1}$
Fiducial(L1_m9)(4)	0	0	2.60	-1.00	-0.28	14.0	$1.92^{+0.5}_{-0.42}$	14.0 ± 0.16	$-1.52^{+0.37}_{-0.41}$	5.00 ± 0.7
Fiducial(L1_m9)(3)	0	0	2.60	-1.00	-0.28	14.0	1.60	13.9 ± 0.19	-1.51 ± 0.39	4.80 ± 0.65
Jet(10)	$0.697^{+2.4}_{-2.7}$	$-0.227^{+1.1}_{-1.1}$	$2.86^{+2.0}_{-0.98}$	$-0.959^{+0.22}_{-0.21}$	-0.305 ± 0.15	$13.5^{+3.8}_{-3.9}$	$1.50^{+0.72}_{-0.55}$	13.6 ± 0.32	-1.89 ± 0.41	$5.25^{+2.6}_{-1.7}$
Jet(8)	0	0	$2.92^{+2.0}_{-1.0}$	-0.956 ± 0.2	-0.313 ± 0.15	$14.0^{+3.6}_{-4.2}$	$1.45^{+0.72}_{-0.58}$	13.7 ± 0.28	-1.86 ± 0.44	$5.01^{+2.1}_{-1.5}$
Jet(6)	0	0	2.60	-1.00	$-0.282^{+0.13}_{-0.13}$	$14.7^{+3.1}_{-4.1}$	$1.45^{+0.76}_{-0.54}$	13.6 ± 0.27	$-1.80^{+0.38}_{-0.41}$	5.98 ± 1.4
Jet(4)	0	0	2.60	-1.00	-0.28	14.0	$1.49^{+0.74}_{-0.5}$	13.6 ± 0.26	$-1.84^{+0.38}_{-0.42}$	6.15 ± 0.98
Jet(3)	0	0	2.60	-1.00	-0.28	14.0	1.60	13.7 ± 0.16	-1.87 ± 0.39	$6.23^{+0.89}_{-0.75}$
Jet_fgas-4σ(10)	$1.79^{+1.9}_{-2.3}$	$-0.785^{+1.2}_{-1.0}$	$2.70^{+1.6}_{-0.77}$	-1.08 ± 0.15	$-0.249^{+0.13}_{-0.23}$	$13.5^{+3.4}_{-3.2}$	1.93 ± 0.39	14.0 ± 0.2	$-1.13^{+0.34}_{-0.37}$	$6.84^{+2.8}_{-2.0}$
Jet_fgas-4σ(8)	0	0	$2.86^{+1.6}_{-0.84}$	-1.07 ± 0.13	$-0.295^{+0.16}_{-0.34}$	$14.3^{+3.2}_{-3.5}$	1.92 ± 0.37	14.1 ± 0.16	$-0.99^{+0.32}_{-0.35}$	$5.78^{+1.9}_{-1.3}$
Jet_fgas-4σ(6)	0	0	2.60	-1.00	$-0.328^{+0.17}_{-0.27}$	$14.9^{+2.8}_{-3.5}$	$1.94^{+0.36}_{-0.32}$	14.1 ± 0.14	$-0.961^{+0.24}_{-0.31}$	6.14 ± 0.98
Jet_fgas-4σ(4)	0	0	2.60	-1.00	-0.28	14.0	$1.96^{+0.35}_{-0.31}$	14.1 ± 0.13	$-1.05^{+0.21}_{-0.26}$	$6.12^{+0.69}_{-0.62}$
Jet_fgas-4σ(3)	0	0	2.60	-1.00	-0.28	14.0	1.60	14.0 ± 0.16	$-1.02^{+0.21}_{-0.25}$	5.94 ± 0.61
fgas$+2\sigma$(10)	$0.761^{+2.4}_{-2.6}$	$-0.328^{+1.1}_{-1.1}$	$2.77^{+2.0}_{-0.9}$	-1.07 ± 0.22	-0.291 ± 0.15	$13.7^{+3.6}_{-4.0}$	$1.58^{+0.71}_{-0.59}$	13.7 ± 0.29	-1.88 ± 0.43	$5.08^{+2.4}_{-1.7}$
fgas$+2\sigma$(8)	0	0	$2.75^{+1.9}_{-0.93}$	$-1.05^{+0.2}_{-0.19}$	-0.302 ± 0.15	$14.1^{+3.4}_{-3.9}$	$1.57^{+0.72}_{-0.59}$	13.7 ± 0.24	-1.85 ± 0.43	$4.82^{+2.1}_{-1.4}$
fgas$+2\sigma$(6)	0	0	2.60	-1.00	$-0.317^{+0.13}_{-0.13}$	$15.0^{+3.2}_{-3.9}$	$1.66^{+0.75}_{-0.64}$	13.7 ± 0.25	-1.80 ± 0.43	$5.19^{+1.1}_{-1.1}$
fgas$+2\sigma$(4)	0	0	2.60	-1.00	-0.28	14.0	$1.66^{+0.79}_{-0.61}$	13.7 ± 0.25	-1.91 ± 0.4	5.10 ± 0.73
fgas$+2\sigma$(3)	0	0	2.60	-1.00	-0.28	14.0	1.60	13.7 ± 0.16	-1.95 ± 0.38	$5.05^{+0.68}_{-0.59}$
fgas-2σ(10)	$1.69^{+2.0}_{-2.2}$	$-0.802^{+1.0}_{-0.95}$	$2.59^{+1.5}_{-0.73}$	-1.06 ± 0.16	$-0.237^{+0.12}_{-0.19}$	$13.8^{+3.5}_{-3.3}$	1.84 ± 0.45	13.9 ± 0.22	$-1.31^{+0.43}_{-0.39}$	$5.53^{+2.3}_{-1.6}$
fgas-2σ(8)	0	0	$2.73^{+1.4}_{-0.88}$	-1.03 ± 0.15	$-0.267^{+0.14}_{-0.29}$	$14.5^{+3.2}_{-3.5}$	$1.83^{+0.45}_{-0.43}$	14.0 ± 0.18	$-1.17^{+0.41}_{-0.43}$	$4.86^{+1.8}_{-1.2}$
fgas-2σ(6)	0	0	2.60	-1.00	$-0.295^{+0.15}_{-0.23}$	$15.2^{+2.8}_{-3.7}$	$1.86^{+0.45}_{-0.38}$	14.0 ± 0.17	$-1.11^{+0.33}_{-0.42}$	$5.08^{+0.83}_{-0.93}$
fgas-2σ(4)	0	0	2.60	-1.00	-0.28	14.0	$1.88^{+0.4}_{-0.36}$	14.0 ± 0.17	$-1.23^{+0.32}_{-0.34}$	4.99 ± 0.63
fgas-2σ(3)	0	0	2.60	-1.00	-0.28	14.0	1.60	14.0 ± 0.22	$-1.20^{+0.34}_{-0.35}$	4.85 ± 0.59
fgas-4σ(10)	$1.68^{+2.0}_{-2.1}$	$-0.822^{+1.0}_{-0.95}$	$2.64^{+1.4}_{-0.73}$	-1.04 ± 0.15	$-0.24^{+0.13}_{-0.24}$	$13.8^{+3.3}_{-3.1}$	1.89 ± 0.41	14.0 ± 0.2	$-1.12^{+0.35}_{-0.37}$	$5.64^{+2.1}_{-1.5}$
fgas-4σ(8)	0	0	$2.81^{+1.4}_{-0.83}$	-1.02 ± 0.13	$-0.293^{+0.17}_{-0.34}$	$14.4^{+3.1}_{-3.4}$	$1.86^{+0.4}_{-0.37}$	14.0 ± 0.16	$-0.964^{+0.31}_{-0.36}$	$4.86^{+1.5}_{-1.0}$
fgas-4σ(6)	0	0	2.60	-1.00	$-0.309^{+0.16}_{-0.25}$	$15.1^{+2.8}_{-3.6}$	$1.85^{+0.39}_{-0.34}$	14.0 ± 0.15	$-0.935^{+0.25}_{-0.32}$	5.23 ± 0.85

Continued on next page

Table 5.4: The best-fit BCM parameters obtained from the joint fitting of all HOS to different FLAMINGO variants. These parameter combinations yield HOS that agree with FLAMINGO measurements to 2%. The integers following the feedback variant names are the number of free parameters. The fixed parameters are highlighted in red. The uncertainty is the 16% and 84% credible interval.

fgas$-4\sigma(4)$	0	0	2.60	-1.00	-0.28	14.0	$1.88^{+0.37}_{-0.31}$	14.0 ± 0.15	$-1.02^{+0.22}_{-0.28}$	5.16 ± 0.58
fgas$-4\sigma(3)$	0	0	2.60	-1.00	-0.28	14.0	1.60	14.0 ± 0.19	$-0.983^{+0.23}_{-0.29}$	5.05 ± 0.54
fgas$-8\sigma(10)$	$1.51^{+2.0}_{-2.1}$	$-0.753^{+1.1}_{-0.93}$	$2.74^{+1.4}_{-0.76}$	-1.04 ± 0.13	$-0.275^{+0.16}_{-0.32}$	14.0 ± 3.2	1.92 ± 0.38	14.0 ± 0.19	$-0.93^{+0.28}_{-0.3}$	$5.94^{+2.4}_{-1.6}$
fgas$-8\sigma(8)$	0	0	$2.96^{+1.4}_{-0.85}$	-1.03 ± 0.12	$-0.323^{+0.18}_{-0.36}$	$14.6^{+3.0}_{-3.4}$	1.93 ± 0.36	14.1 ± 0.14	$-0.857^{+0.26}_{-0.24}$	$5.12^{+1.5}_{-1.0}$
fgas$-8\sigma(6)$	0	0	2.60	-1.00	$-0.322^{+0.17}_{-0.25}$	$15.0^{+2.7}_{-3.4}$	$1.91^{+0.35}_{-0.33}$	14.1 ± 0.12	$-0.848^{+0.21}_{-0.2}$	5.63 ± 0.83
fgas$-8\sigma(4)$	0	0	2.60	-1.00	-0.28	14.0	1.92 ± 0.35	14.1 ± 0.12	$-0.893^{+0.16}_{-0.19}$	5.57 ± 0.56
fgas$-8\sigma(3)$	0	0	2.60	-1.00	-0.28	14.0	1.60	14.0 ± 0.14	$-0.885^{+0.17}_{-0.18}$	5.40 ± 0.52
M*$-\sigma(10)$	$1.70^{+2.0}_{-2.3}$	$-0.901^{+1.1}_{-0.94}$	$2.58^{+1.5}_{-0.74}$	-1.06 ± 0.16	$-0.232^{+0.12}_{-0.22}$	$13.6^{+3.5}_{-3.3}$	$1.89^{+0.47}_{-0.44}$	13.9 ± 0.22	$-1.28^{+0.43}_{-0.4}$	$5.48^{+2.2}_{-1.6}$
M*$-\sigma(8)$	0	0	$2.75^{+1.4}_{-0.87}$	-1.04 ± 0.15	$-0.263^{+0.14}_{-0.29}$	$14.4^{+3.2}_{-3.4}$	$1.85^{+0.48}_{-0.42}$	14.0 ± 0.18	$-1.15^{+0.4}_{-0.44}$	$4.73^{+1.6}_{-1.1}$
M*$-\sigma(6)$	0	0	2.60	-1.00	$-0.296^{+0.15}_{-0.25}$	$15.0^{+2.9}_{-3.7}$	$1.90^{+0.46}_{-0.39}$	14.0 ± 0.16	$-1.09^{+0.35}_{-0.42}$	$4.94^{+0.81}_{-0.92}$
M*$-\sigma(4)$	0	0	2.60	-1.00	-0.28	14.0	$1.92^{+0.42}_{-0.36}$	14.0 ± 0.16	$-1.20^{+0.32}_{-0.36}$	4.85 ± 0.57
M*$-\sigma(3)$	0	0	2.60	-1.00	-0.28	14.0	1.60	14.0 ± 0.21	$-1.17^{+0.33}_{-0.37}$	4.69 ± 0.55
M*$-\sigma$_fgas$-4\sigma(10)$	$1.60^{+2.0}_{-2.2}$	$-0.863^{+1.0}_{-0.9}$	$2.61^{+1.3}_{-0.73}$	-1.02 ± 0.13	$-0.268^{+0.15}_{-0.34}$	13.7 ± 3.3	$1.92^{+0.41}_{-0.36}$	14.0 ± 0.19	$-0.906^{+0.29}_{-0.31}$	$5.64^{+2.2}_{-1.5}$
M*$-\sigma$_fgas$-4\sigma(8)$	0	0	$2.89^{+1.3}_{-0.86}$	-1.01 ± 0.12	$-0.329^{+0.19}_{-0.41}$	$14.4^{+3.0}_{-3.3}$	$1.91^{+0.39}_{-0.35}$	14.0 ± 0.14	-0.814 ± 0.27	$4.81^{+1.4}_{-0.97}$
M*$-\sigma$_fgas$-4\sigma(6)$	0	0	2.60	-1.00	$-0.33^{+0.17}_{-0.27}$	$15.1^{+2.6}_{-3.4}$	$1.88^{+0.38}_{-0.34}$	14.0 ± 0.13	-0.809 ± 0.23	$5.28^{+0.72}_{-0.81}$
M*$-\sigma$_fgas$-4\sigma(4)$	0	0	2.60	-1.00	-0.28	14.0	$1.91^{+0.36}_{-0.3}$	14.0 ± 0.12	-0.867 ± 0.2	5.16 ± 0.51
M*$-\sigma$_fgas$-4\sigma(3)$	0	0	2.60	-1.00	-0.28	14.0	1.60	14.0 ± 0.15	-0.852 ± 0.2	5.05 ± 0.5

Baryonification error for strong and weak feedback variants

Following the discussion in § 5.7, we show here the residual errors for the best-fit BCM models applied to two extreme feedback variants from the FLAMINGO suite: the strongest (fgas-8σ) and the weakest (fgas$+2\sigma$). These figures are analogous to Fig. 5.5, which showed results for the fiducial case. For the strong feedback case (fgas-8σ), the BCM reproduces most statistics to within 2%, with a mild 3–4% excess in m_4 at low redshift and small angular scales (Fig. 5.9). This reflects both modeling limitations and the intrinsic noise in higher-order moments. For the weak feedback case (fgas$+2\sigma$), residuals remain within 2% across all statistics and scales (Fig. 5.10).

Cosmology dependence of baryonic suppression

To test how baryonic suppression of weak lensing summary statistics varies with cosmology, we compute the statistics for both the fiducial and LS8 cosmologies, and for each with the fiducial and strong-AGN (fgas-8σ) feedback variants. Fig. 5.11 shows the relative difference in suppression between LS8 and fiducial cosmologies in the first redshift bin, which is sub-percent for all statistics and scales.

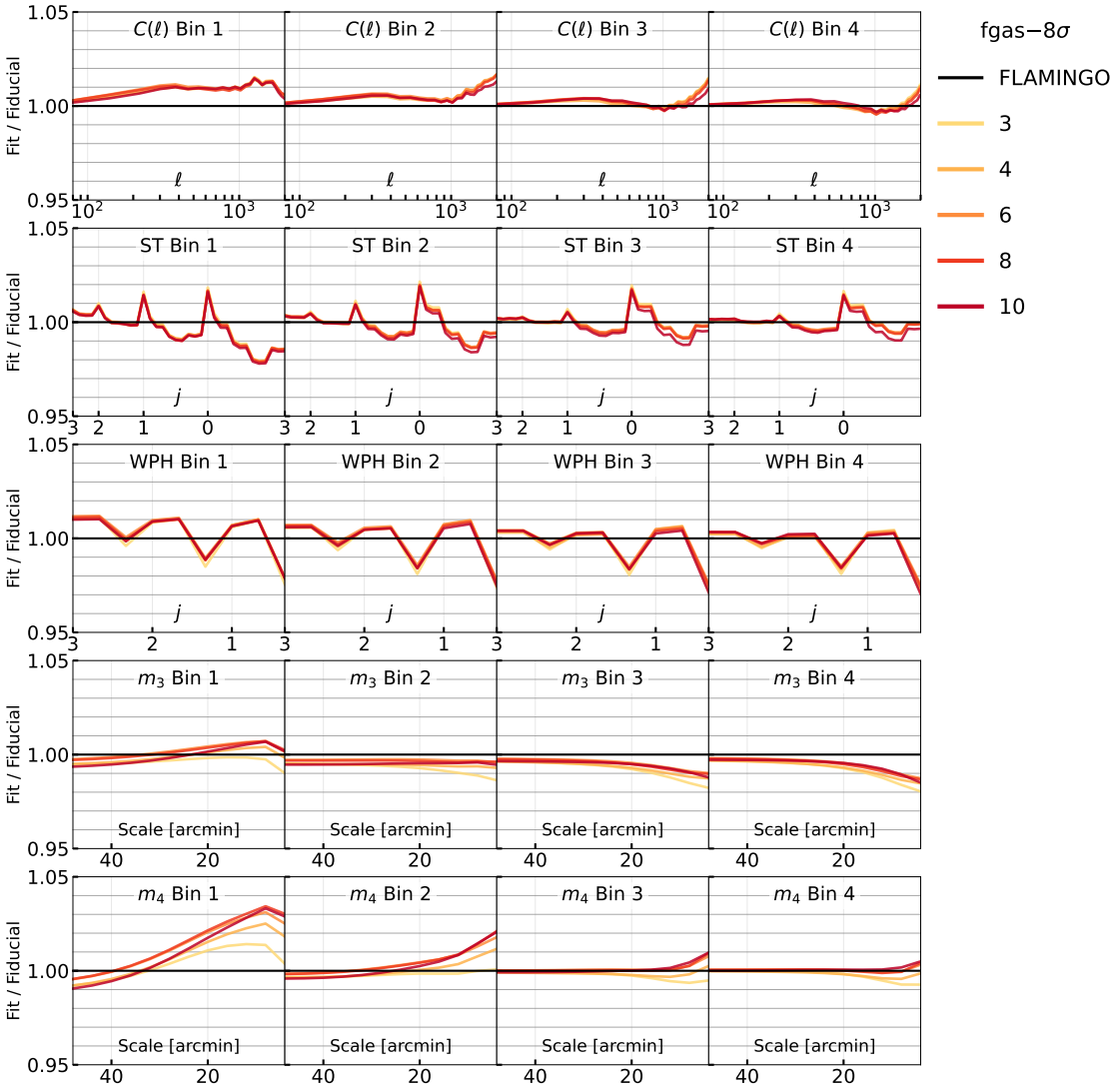

Figure 5.9: Similar to Fig. 5.5 but with BCM fitted to the strongest feedback variant in the FLAMINGO suite. The BCM error is still at the percent level. The elevated residual for low redshift m_4 is potentially related to both the baryon modelling error and/or m_4's sensitivity to noise.

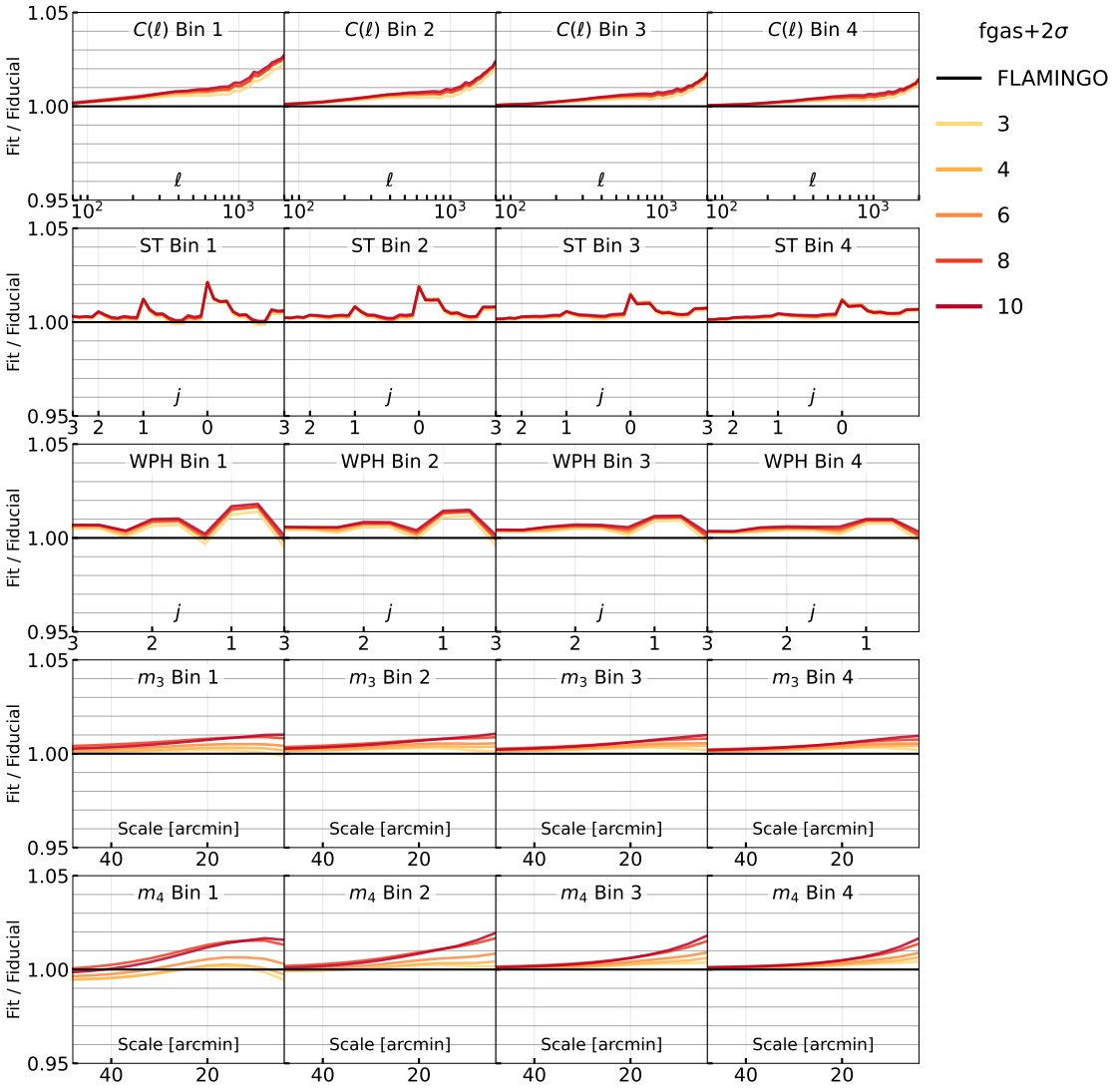

Figure 5.10: Similar to Fig. 5.5 but with BCM fitted to the weakest feedback variant in the FLAMINGO suite. The BCM error is still at the percent level.

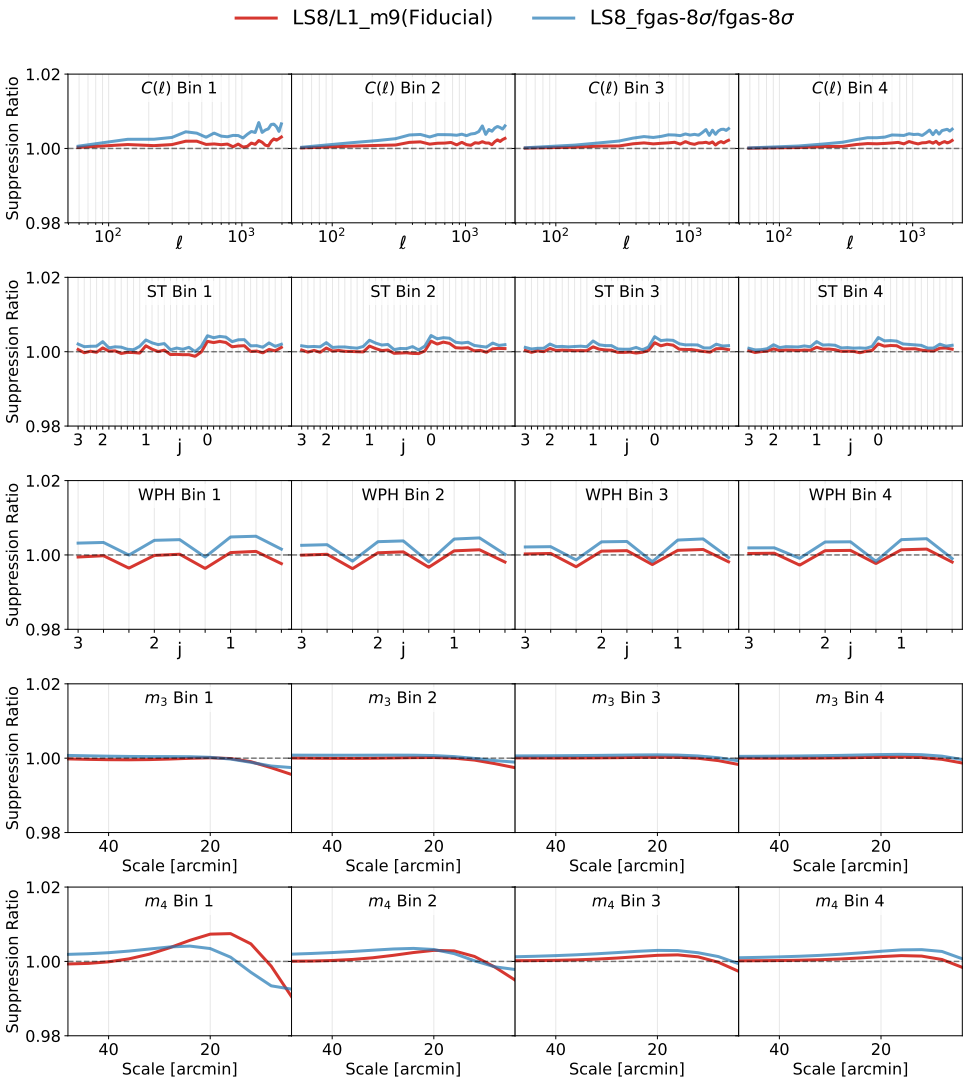

Figure 5.11: The sensitivity of the baryonic effect to cosmology for the first redshift bin. The y-axis shows the fractional change between the baryonic suppression of summary statistics at the LS8 cosmology compared to the suppression at the fiducial cosmology, for two different subgrid models: the fiducial models (LS8 and L1_m9) and the strong AGN feedback variants (LS8_fgas-8σ and fgas-8σ). Across both feedback variants, the amount of baryonic suppression changes only at the sub-percent level, indicating weak dependence on σ_8.

CHAPTER 6

THE ISOTROPY OF HUBBLE EXPANSION IN THE EARLY AND LATE UNIVERSE

6.1 Abstract

We test the isotropy of Hubble expansion by combining several probes for the first time, constructing full-sky maps of expansion rate variation using Type Ia supernovae, fundamental plane galaxies, and CMB temperature fluctuations. We find no hint of anisotropy or correlation between early- and late-Universe expansion across all systematic models. The 99% confidence upper limits on expansion rate anisotropy are 0.39% for low-redshift supernovae, 0.95% for high-redshift CMB, and 0.37% when combined at a 60-degree smoothing scale. A significant anomaly in the fundamental plane residual map may reflect systematics in the current DESI dataset, as evidenced by the absence of cross-correlation with other tracers and its correlation with spatial density variations.

6.2 Introduction

The standard ΛCDM cosmological model assumes homogeneity and isotropy: that the laws of physics are invariant throughout the Universe and that the distribution of matter and energy is statistically uniform on large scales. A direct consequence is an isotropic cosmic expansion. The expansion history, $H(z)$, and its present-day value, the Hubble constant $H_0 = 100h$ km s^{-1} Mpc^{-1}, are constrained by early-Universe

temperature anisotropies of the Cosmic Microwave Background (CMB) and by late-Universe distance indicators such as Type Ia supernovae (SNe) and fundamental plane (FP) galaxies.

Yet two decades of CMB observations [284–291] have revealed hints of statistical anomalies in the low-ℓ temperature multipoles. Low-redshift surveys have also raised questions about the nature of the recent cosmic acceleration [292, 293]. Furthermore, h inferred from early-Universe CMB data remains in tension with that from local supernovae measurements [294, 295]. A non-uniform Hubble flow could potentially alleviate these tensions through scenarios such as vector field inflation, non-trivial topologies, or inhomogeneous dark energy [286, 296–298]. However, the scarcity of independent large-scale modes makes it difficult to quantify such anomalies precisely.

We test for anisotropy in the Hubble flow by combining information from both early- and late-Universe probes for the first time. Under ΛCDM, any directional variation in H_0 measurements should be *consistent with statistical noise and sample variance* and also *uncorrelated between epochs*. A significant violation of this null hypothesis would suggest new physics or previously unaccounted-for systematics. While Hubble flow isotropy has been examined in individual probes [299–307], we systematically construct full-sky h maps using SNe, FP, and CMB data and compare them with isotropic predictions. The SN and FP maps trace the local expansion ($z < 1$), whereas the CMB map extrapolates the expected h from the $z \approx 1100$ temperature fluctuations. Not only do they independently probe the isotropy of the low- and high-z Hubble flow, but their cross-correlation also provides a robust test for any correlated signal between the early and late Universe.

6.3 Hubble residual maps

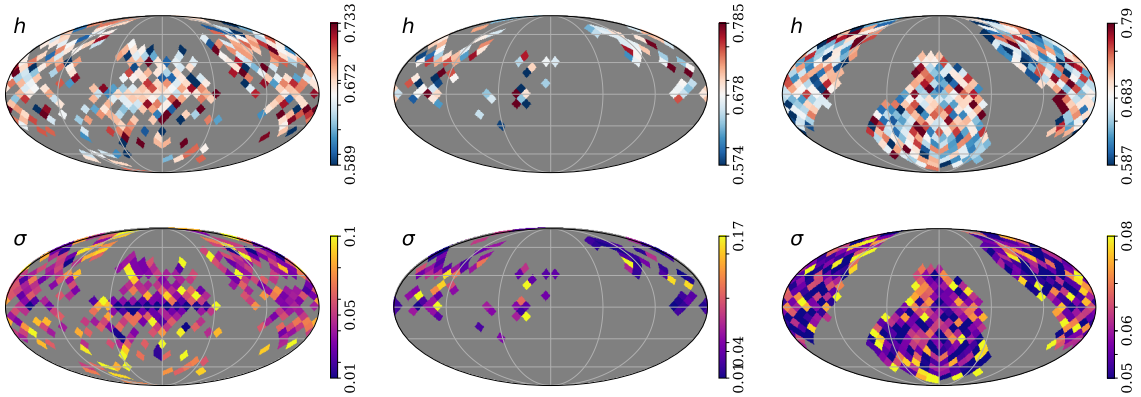

Figure 6.1: The h and uncertainty σ maps for **supernovae** (left), **fundamental plane galaxies** (middle) and **CMB** (right). All maps are calibrated to the Planck cosmology using fiducial analysis settings. Color bars mark the 5th, 50th and 95th percentiles. $f_{\rm sky} = 0.41, 0.12, 0.52$.

Supernovae. Type Ia supernovae (SNe) are precise standardizable candles for cosmological distance measurements. The primary observable is the distance modulus $\mu = m - M$, where m and M are the apparent and standardized absolute magnitudes, respectively. For a supernova i at scale factor a_i (redshift z_i), the theoretical distance modulus $\mu_{{\rm th},i}$ for a flat universe is

$$\mu_{{\rm th},i} = 5\log_{10}\left[\frac{1}{\rm Mpc}\left(\frac{c}{a_i}\int_{a_i}^{1}\frac{da'}{a'^2 H(a')}\right)\right] + 25\,, \tag{6.1}$$

where the argument of the logarithm is the luminosity distance d_L in Mpc. Equation 6.1 assumes an isotropic Hubble flow. We hypothesize that the difference between the observed μ_i and $\mu_{{\rm th},i}$ arises from noise and potential anisotropy in the cosmic expansion. We model the latter as a perturbation to the Hubble parameter

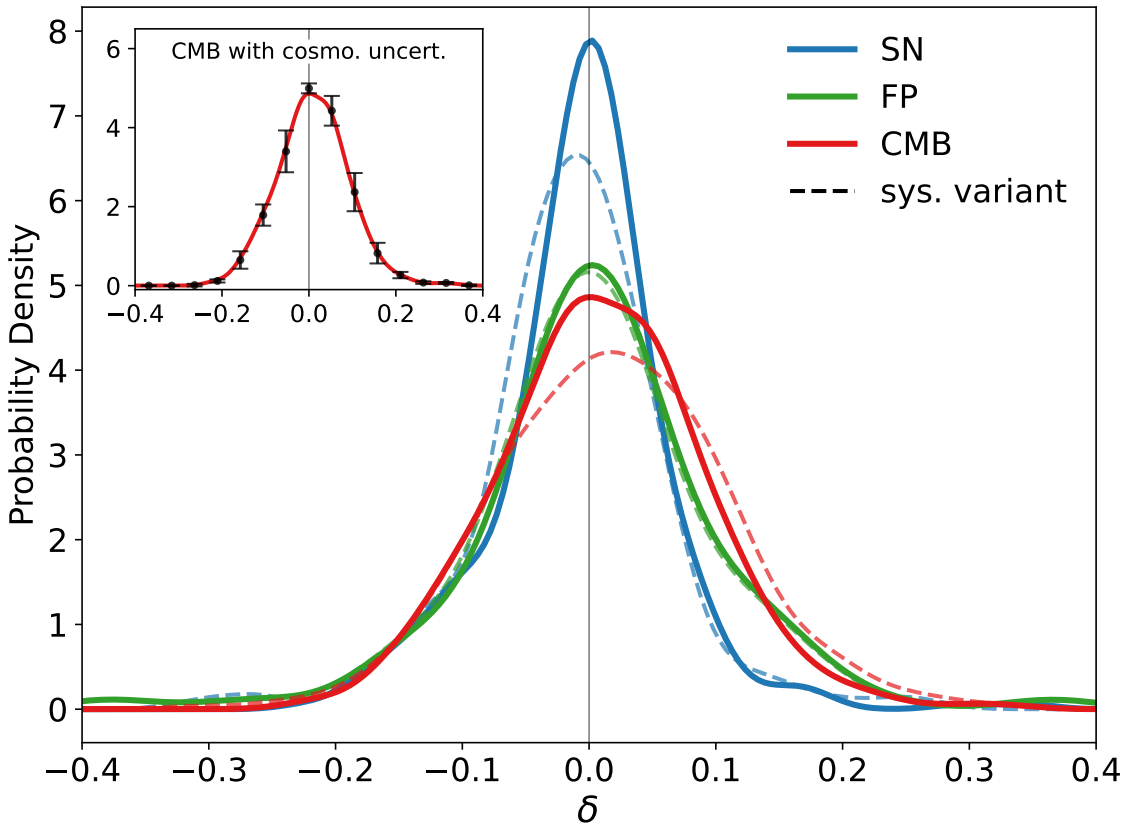

Figure 6.2: Pixel value histograms of $\delta_{\rm SN}$, $\delta_{\rm FP}$, and $\delta_{\rm CMB}$. Solid lines show fiducial configurations; dashed lines show systematic variants. The inset shows $\delta_{\rm CMB}$ histograms with error bars indicating 1σ cosmological uncertainty. The distributions are zero-centered and approximately Gaussian. $\delta_{\rm SN}$ has the smallest scatter and $\delta_{\rm CMB}$ the largest.

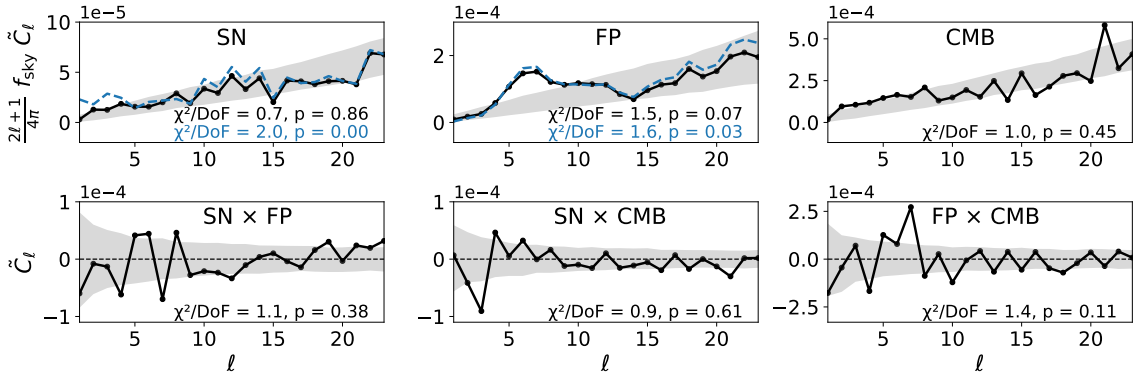

Figure 6.3: Weighted C_ℓ (Eq. 6.13) of the Hubble residual maps $\delta_{\rm SN}$, $\delta_{\rm FP}$, and $\delta_{\rm CMB}$. The weighted auto-C_ℓ is scaled by $f_{\rm sky}$ so that the area under the curve approximates the variance of the full-sky δ field. Solid lines represent the data, while shaded regions represent the 1σ intervals of isotropic noise realizations. For the fiducial model (black), $\delta_{\rm SN}$ and $\delta_{\rm CMB}$ are consistent with isotropy, whereas the FP auto-C_ℓ is not, with a significant peak around $\ell = 6$. The SN and FP systematic variants (blue) do not remove peculiar velocities and hence exhibit anisotropic contamination from the large-scale structures.

along the line of sight to each supernova

$$H_{0,i} = H_0(1 + \delta_i)\,. \tag{6.2}$$

In the limit of noiseless measurement, the distance modulus residual is directly proportional to the Hubble flow perturbation [302], since

$$\mu_i - \mu_{{\rm th},i} = 5\log_{10}\left(\int_{a_i}^1 \frac{da'}{a'^2 H_i(a')} \Big/ \int_{a_i}^1 \frac{da'}{a'^2 H(a')}\right) \tag{6.3}$$

$$= -5\log_{10}(1+\delta_i) \approx -\frac{5\,\delta_i}{\ln 10}\,. \tag{6.4}$$

Equation 6.4 allows us to estimate δ_i and its uncertainty σ_i from each supernova.

Next, we directly construct a full sky δ map using the `nside` $= 8$ HEALPix pixelization scheme [54, 308] in contrast to the more indirect normalized residual approach of Soltis *et al.* [302]. The value and uncertainty of each pixel p is given by the unbiased inverse variance estimator

$$\delta_{\rm SN}(p) = \frac{\sum_{i\in p} \delta_i/\sigma_i^2}{\sum_{i\in p} 1/\sigma_i^2}\,, \quad \sigma_{\rm SN}(p) = \frac{1}{\sqrt{\sum_{i\in p} 1/\sigma_i^2}}\,. \tag{6.5}$$

We analyze the Pantheon+SH0ES SNe data from Riess *et al.* [309], originally calibrated via Cepheids to the absolute magnitude $M_{\rm fid} = -19.24$ and yielding a global measurement of $h = 0.7330$. Since we focus on the anisotropy rather than the absolute value of h and wish to compare with other distance tracers, we recalibrate the SNe to the Planck 2018 CMB cosmology ($h = 0.6736$ [14]), thereby changing the absolute magnitude to $M = -19.42$. The recalibration only centers the distribution of h around $h_{\rm Planck}$ but has no impact on our anisotropy results. We include 1701 SN with $z < 2.3$ and median $z = 0.164$. The h and uncertainty maps are shown in Fig. 6.1, and the pixel histogram in Fig. 6.2. The relative motion between the SNe and the observer modifies the SNe redshift and coherently distorts of $\delta_{\rm SN}$ [310, 311]. In the fiducial analysis, we consider supernova redshifts corrected for both the CMB dipole and peculiar velocities due to local group and host motions [309]. The systematic variant where the redshift ($z_{\rm CMB}$) is corrected only for the former is shown in Fig. 2 as the blue dashed curve.

Fundamental Plane Galaxies. Empirically, the physical radius of elliptical galaxies is related to their dynamical properties, and can thus be calibrated as a standard-

ized ruler for distance and expansion rate measurements. This relationship, called the fundamental plane (FP), relates a galaxy's effective radius R_e, central velocity dispersion σ, and mean surface brightness within R_e, I_e,

$$\log R_e = a \log \sigma + b I_e + c \,. \tag{6.6}$$

We analyze the DESI FP galaxy data [312], where a and b are calibrated to the observed galaxy population, and the zero-point c is calibrated to the surface brightness fluctuation distance of the Coma cluster [312, 313]. We estimate each galaxy's angular diameter distance d_A (Mpc) by combining the FP-inferred physical radius R_e (kpc) with the observed angular size θ_e (arcseconds),

$$k_i \equiv \log_{10} d_{A,i} = \log_{10} \left(\frac{R_{e,i}}{\theta_{e,i}} \cdot \frac{648}{\pi} \right) , \tag{6.7}$$

where $d_{A,i}$ is the angular diameter distance .

Analogous to the SNe analysis, if the expansion rate in direction i deviates from the homogeneous background by a perturbation δ_i (Eq. 6.2), the observed angular diameter distance differs from the isotropic theory by

$$k_i - k_{\rm th,i} = \log_{10} \left(\int_{a_i}^{1} \frac{da'}{a'^2 H_i(a')} \Big/ \int_{a_i}^{1} \frac{da'}{a'^2 H(a')} \right) \tag{6.8}$$

$$= \log_{10} \left(\frac{1}{1+\delta_i} \right) \approx -\frac{\delta_i}{\ln 10} \,. \tag{6.9}$$

Similar to the SNe, this distance measurement is also degenerate with h. Using $h = 0.7605$ from the DESI analysis, we recalibrate k_i to the Planck cosmology by

$k_i \to k_i + \log_{10}(h_{\rm DESI}/h_{\rm Planck})$. The recalibration centers the distribution of h around $h_{\rm Planck}$.

We consider 4063 galaxies with $0.023 < z < 0.1$ and median $z = 0.075$. We construct $\delta_{\rm FP}$ and $\sigma_{\rm FP}$ maps by inverse-variance averaging the FP galaxy measurements on a HEALPix grid (analogous to Eq. 6.5). The maps and the pixel histograms are shown in Figs. 6.1 and 6.2. The fiducial analysis corrects the redshift with the 2M++ peculiar velocity (PV) model [314], while a systematic variant does not.

The Cosmic Microwave Background. The CMB temperature power spectrum C_ℓ is another sensitive probe of h [5, 14]. To construct the Hubble residual map, we partition the CMB temperature map into 768 HEALPix patches (labeled by p), measure the local power spectrum $C_\ell^{\rm local}$, and finally estimate the residual $\delta_{\rm CMB}(p) = h(p)/h_{\rm Planck} - 1$ through a likelihood analysis.

We use the two Planck 2018 half-mission SMICA maps [8, 14], T_1 and T_2 with a resolution of `nside` = 1024. We use the combined 100, 143, and 217 GHz frequency masks ($f_{\rm sky} \approx 0.5$) to mitigate galactic foreground and point source contamination. For each pixel p, we gnomonically project T_1, T_2, and the masks to a Cartesian patch. We apodize the mask, measure the mode-decoupled C_ℓ, and compute the Gaussian covariance of C_ℓ, Σ, using `pymaster` [315]. Since the two half-mission temperature maps measure the same CMB structures but independent noise realizations, their cross-spectrum directly estimates the temperature auto-spectrum without noise bias.

For each sky patch, we measured the local power spectrum $C_\ell^{\rm obs}$, and estimate the local h using a maximum-likelihood approach using the two-point statistics. The

theoretical model for $C_\ell^{\rm obs}$ is

$$C_\ell^{\rm local} = C_\ell^{\rm th} \cdot P_{\ell,1024}^2 \cdot P_{\ell,\rm proj}^2 \cdot B_\ell^2 \cdot T_{\ell,\rm SMICA}, \tag{6.10}$$

where $C_\ell^{\rm th}$ is the full-sky LCDM prediction by `camb` [99], P_{1024} and $P_{\rm proj}$ are the pixel window functions of the HEALPix and the Cartesian projection, B and $T_{\rm SMICA}$ are the SMICA beam and transfer function [8]. Since we are interested in the Hubble flow and not the CMB physics, we fix all other parameters (including $\Omega_m h^2, \Omega_b h^2$) and vary only h as we compute the posterior

$$\log \mathcal{P}(h) \propto \sum_{\ell,\ell'} \left(C_\ell^{\rm obs} - C_\ell^{\rm local}(h)\right) \Sigma_{\ell\ell'}^{-1} \left(C_{\ell'}^{\rm obs} - C_{\ell'}^{\rm local}(h)\right). \tag{6.11}$$

We construct the residual maps $\delta_{\rm CMB}(p) = h(p)/h_{\rm Planck} - 1$ and $\sigma_{\rm CMB} = \sigma(p)/h_{\rm Planck}$. The maps and the pixel histograms are shown in Figs. 6.1 and 6.2.

In the fiducial analysis, we consider the scale cut $\ell \in [100, 1250]$ in the likelihood analysis. A more conservative systematic analysis uses a more aggressive scale-cut $\ell \in [200, 1000]$. Most of the h sensitivity lies near the second acoustic peak at $\ell \approx 500$, which falls within both scale cuts. Larger scales exceed the patch size and are noisy; smaller scales are removed due to potential point source contamination. We also assess the impact of cosmological uncertainty in $\delta_{\rm CMB}$ by creating 1000 maps with cosmology sampled from the Planck posterior [14].

6.4 Analysis

We jointly analyze the statistics of the three Hubble residual maps: $\delta_{\rm SN}$, $\delta_{\rm FP}$, and $\delta_{\rm CMB}$. To account for incomplete sky coverage and spatially varying measurement uncertainties, we focus on an inverse-variance weighted version of the C_ℓ statistics [302, 315],

$$\tilde{a}^{\alpha}_{\ell m} = \frac{\int M(\hat{\mathbf{n}})\delta_\alpha(\hat{\mathbf{n}})\sigma_\alpha^{-2}(\hat{\mathbf{n}})Y^*_{\ell m}(\hat{\mathbf{n}})\,d\Omega(\hat{\mathbf{n}})}{\int M(\hat{\mathbf{n}})\sigma_\alpha^{-2}(\hat{\mathbf{n}})\,d\Omega(\hat{\mathbf{n}})/4\pi}\,, \tag{6.12}$$

$$\tilde{C}^{\alpha,\beta}_\ell = \langle \tilde{a}^{\alpha}_{\ell m}\tilde{a}^{\beta,*}_{\ell m}\rangle_m\,, \tag{6.13}$$

where $\ell \in [1, 23]$, α, β refer to SNe, FP, or CMB, and $M(\hat{\mathbf{n}})$ is the mask. Eqs. 6.12 and 6.13 are closely related to the standard C_ℓ statistics – if the unmasked data have spatially constant uncertainty and a flat power spectrum, Eq. 6.12 reduces to $\tilde{a}_{\ell m} = f_{\rm sky}^{-1}\int M(\hat{\mathbf{n}})\delta(\hat{\mathbf{n}})Y^*_{\ell m}(\hat{\mathbf{n}})\,d\Omega(\hat{\mathbf{n}})$, and the weighted power spectrum is related to the standard one by

$$C_\ell \approx f_{\rm sky}\tilde{C}_\ell\,. \tag{6.14}$$

To test for isotropy, we extend the approach of Soltis *et al.* [302] to multiple tracers. Consider the set $\{(\delta(p), \sigma(p))\}$ for all *unmasked* pixels p. This *distribution* includes contributions from both the potentially anisotropic Hubble field and the underlying noise properties. For example, it captures how the spatially varying mask fraction of the CMB affects $\delta_{\rm CMB}$ and $\sigma_{\rm CMB}$, and how individual object measurement uncertainties and spatial number density variations affect the SNe and FP maps. For each δ map, we create 1000 *isotropic noise realizations* by drawing unmasked pixel

values from this distribution $\{(\delta(p), \sigma(p))\}$. For each realization, we compute the auto- and cross-power spectra $\tilde{C}_\ell^{\mathrm{iso},\alpha\beta}$ among all tracers. Fig. 6.3 compares $\tilde{C}_\ell$ and $\tilde{C}_\ell^{\mathrm{iso},\alpha\beta}$ for the fiducial (black) and the systematics models (blue).

To quantitatively evaluate how well each tracer individually, as well as their combination, supports or rejects the isotropy hypothesis, we perform a χ^2 test on the auto- and cross-spectra. We consider the multipole range $\ell \in [1, 23]$ and estimate the covariance matrix from the isotropic noise realizations. To account for finite sample bias in the inverse covariance estimator, we apply the correction [316, 317]

$$\chi^2 = \chi^2_{\mathrm{raw}} \times \frac{N_{\mathrm{sim}} - N_{\mathrm{data}} - 2}{N_{\mathrm{sim}} - 1}\,, \tag{6.15}$$

where N_{sim} is the number of isotropic simulations and N_{data} is the length of the data vector. Fig. 6.3 shows the test statistics of the individual auto- and cross-correlations.

Configuration			χ^2/DoF	p-value
SN + CMB *(DoF=66)*		**Fiducial**	**0.92**	**0.66**
	sys. variants	SN: z_{CMB}	1.19	0.14
		CMB: scale-cut	0.96	0.57
		CMB: cosmo. uncert.	0.88	0.75
SN+FP+CMB *(DoF=132)*		Fiducial	1.02	0.41
		FP: no PV	1.21	0.05

Table 6.1: Results of the isotropy χ^2 test using SNe, CMB, and FP tracers. The main result uses only the SNe and the CMB data (row 1), and supports the null hypothesis of isotropy. Lower rows show the test statistics when varying a single tracer's systematic model or combining with the FP data.

The auto and cross-spectra of the fiducial δ_{SN} and δ_{CMB} maps are consistent with isotropy. Their combined statistics have a p-value of 0.66 (Tab. 6.1, row 1) and sup-

ports the null hypotheses of *(1) isotropic expansion and (2) uncorrelated expansion across epochs.* This is the main finding of this letter. The result only holds after SNe peculiar velocity correction, and is robust against conservative CMB scale cuts ($p = 0.57$) and cosmology uncertainty marginalization ($p = 0.75$).

For $\delta_{\rm FP}$, we find a significant excess of power around $\ell = 6$ ($\theta \approx 30^\circ$, $p = 0.07$). Unlike the SNe, the PV correction only alleviates, but cannot eliminate, this anomaly. However, since there is no significant cross-correlation between $\delta_{\rm FP}$ and $\delta_{\rm SN}$ (which cover the same low-redshift volume) and $\delta_{\rm CMB}$, we are reluctant to associate the FP anomaly with true cosmic anisotropy. Another reason against this association is that we find a significant correlation between $\delta_{\rm FP}$ and the FP number density distribution (and to a lesser degree the mean redshift in each pixel) in the same ℓ range. This evidence for a possible systematic effect in the FP sample is robust across all aforementioned systematic choices. The current FP sample comprises only targets from the DESI science validation phase and has significant spatial depth variation across the sky, potentially inducing unaccounted-for selection effects. Future DESI data expects a more homogeneous coverage [312, 318].

To obtain an upper limit on the variance of the anisotropy, we follow the general approach of Soltis *et al.* [302] and decompose the observed δ into a true signal and noise, $\delta_{\rm obs} = \delta_{\rm true} + n$. Since $\delta_{\rm obs}$ for both SNe and CMB are consistent with isotropy, the noise must dominate $\delta_{\rm obs}$, thereby constraining the amplitude of $\delta_{\rm true}$. We construct this limit using both probes via the full-sky power spectrum. Since

$\delta_{\rm obs} \approx n$, the 99% upper bound on the variance of the true δ

$$\left\langle \delta^2_{\rm true} \right\rangle_{99\%} \; < \; \sum_\ell \frac{2\ell+1}{4\pi} \Big(C_\ell^{\rm iso} + 2.33\, \Delta C_\ell^{\rm iso} \Big)\, B_\ell^2\, d\ell \,, \tag{6.16}$$

where $C_\ell^{\rm iso}$ and $\Delta C_\ell^{\rm iso}$ are the mean and scatter of the power spectrum of the full-sky isotropic noise, and B_ℓ smooths the map via a Gaussian beam. The factor 2.33 corresponds to the one-sided 99% confidence threshold for a Gaussian distribution. We derive this limit for the SNe, CMB, and inverse-variance combined maps on their joint footprint for scales between 10 and 120 degrees in Fig. 6.4. At 60 degrees, the 99% confidence limits on low-z and high-z anisotropy are 0.39% and 0.95% respectively, and 0.37% when combined.

6.5 Conclusion

We tested the isotropy of cosmic expansion using three independent distance tracers: Type Ia supernovae, fundamental plane galaxies, and CMB temperature fluctuations. Aside from a localized anomaly in the FP auto-power spectrum around $\ell = 6$, all measurements are consistent with isotropic expansion at low and high redshift, and results are robust across the systematic modeling choices we tested. The SNe and CMB measurements yield stringent constraints on potential Hubble flow anisotropy, with 99% confidence upper limits of 0.39% for low-redshift SNe, 0.95% for high-redshift CMB, and 0.37% when combined at a 60-degree smoothing scale.

We argue that – given this other evidence – it is premature to claim that the FP anomaly is a sign of genuine cosmic anisotropy. This hesitance is supported by the

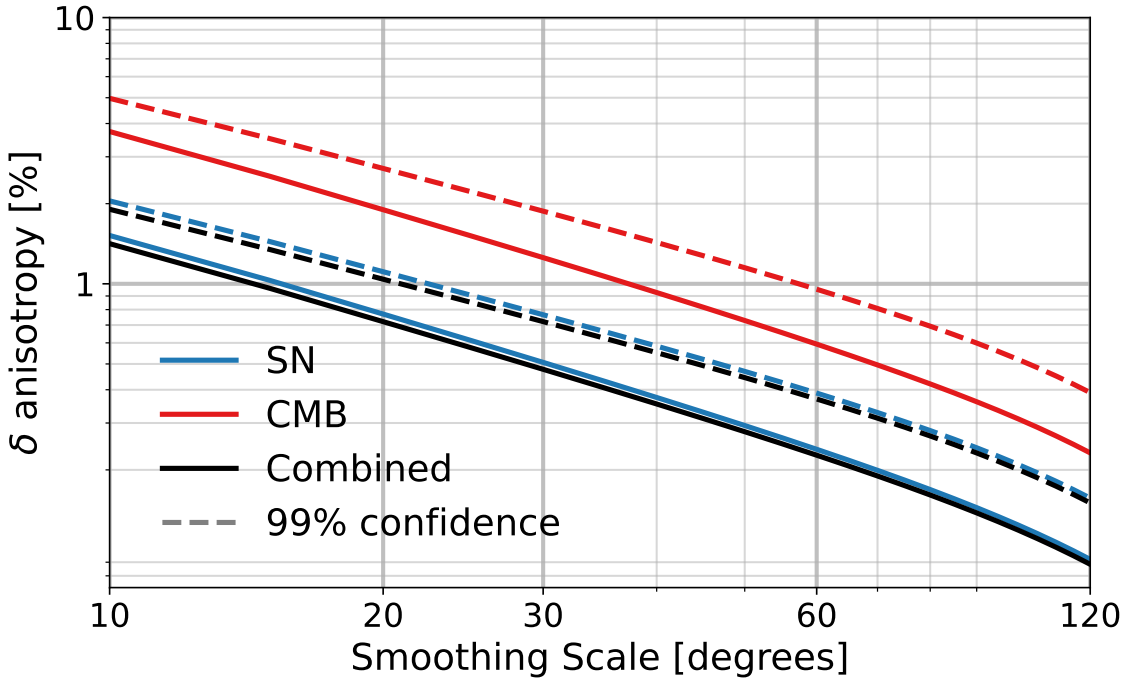

Figure 6.4: Upper limits on Hubble constant anisotropy as a function of smoothing scale. Solid lines show the mean noise level from isotropic realizations, while dashed lines indicate 99% confidence upper limits for the SNe, CMB, and the combined maps.

absence of significant cross-correlation between $\delta_{\rm FP}$ and the other tracers, particularly with $\delta_{\rm SN}$ which probes the same low-redshift volume, and by the correlation between $\delta_{\rm FP}$ and the spatial distribution of FP galaxy density. Future DESI data releases with more homogeneous sky coverage should shed light on this anomaly.

Our results agree with and extend those of Refs. [300, 302]: multiple probes establish that any small Hubble residuals are uncorrelated over low and high redshifts. Our results are complimentary to those of Refs. [304, 319], which focus on low-redshift anisotropy on the largest scales with particular shapes: we jointly studied both early and late-Universe anisotropy, and only focused on the power spectrum.

CHAPTER 7

FIELD-LEVEL DE-CONTAMINATION OF ATMOSPHERIC FOREGROUNDS FOR GROUND-BASED CMB EXPERIMENTS

7.1 Introduction

Observations of the cosmic microwave background (CMB) underpin much of our understanding of the universe. Primarily because of the CMB, we have a fiducial cosmological model (ΛCDM) containing 7 parameters ($\Omega_b, \Omega_c, h, A_s, N, \sum m_\nu, \tau$) that describes almost all data from redshift 1100 until today. Nonetheless, key questions remain: the energy scale of inflation, the evolution of dark energy, the sum of the neutrino masses, and the physics of large-scale structure. A major limiting factor is that large-scale CMB temperature and polarization from the ground remain noise dominated, motivating improved map-level reconstructions.

Ground-based CMB observations are contaminated by atmospheric emission, which can exceed the CMB signal by orders of magnitude at millimeter wavelengths. It is helpful to keep some numbers in mind to understand the scope of the problem. The root-mean-square (RMS) temperature fluctuations in the CMB are of order $80\,\mu$K, polarization E-mode fluctuations an order of magnitude lower than that, and primordial B-modes likely lie below the 100 nK level. Current ground-based experiments scan the sky with thousands of detectors[1], each of which has RMS noise

1. Numbers used throughout the text are close to those in the South Pole Telescope third-generation camera (SPT-3G) unless otherwise specified.

of order $700\,\mu\mathrm{K\,Hz}^{-1/2}$, so a detector that samples at 150 Hz will have RMS noise of order $10^4\,\mu$K in each sample. However, detector noise can be reduced by scanning the same region of the sky with multiple independent detectors many times. As a simple example, if 500 detectors stare at the same 1° patch of the sky for 1 second and record 150 samples each, then the coadded data have detector noise of $10^4\,\mu\mathrm{K}/\sqrt{500\times150} \simeq 40\,\mu$K, of order the CMB RMS. This is for a single stare; thousands of stares render the detector noise smaller still by a factor of 30. This is the reason why ground-based experiments have been able to make cosmic variance limited measurements of the spectrum.

Why then is it difficult for ground-based experiments to measure the CMB on large scales? The problem is emission from poorly mixed water vapor (i.e., clouds) in the atmosphere, emission characterized by a Kolmogorov spectrum falling approximately as $\ell^{-11/3}$ (or even faster) on the largest scales [320, 321]. Estimating the RMS fluctuations to be of order 10 mK at $\ell = 6$ implies that they drop below the signal only at $\ell \sim 500$. To suppress the atmosphere and mitigate $1/f$ noise, experiments (at least at the South Pole) typically apply a high-pass filter, which effectively removes the large-scale CMB.

If the atmosphere were stable over a single short scan and was independent from scan to scan, then the factor of 30 suppression from multiple scans still leaves atmospheric noise larger than the CMB on large scales. The atmosphere, however, is not constant on the sky even within a single scan. Under the frozen flow approximation, the "frozen atmosphere" is blown by the wind [320–325]. The large spatial scale features of the atmospheric emission are retained and simply translated in time. The

two dominant sources of emission then in many ground-based millimeter experiments – the CMB and the atmosphere – have different temporal properties: the atmosphere moves and the CMB does not. We propose to capitalize on this difference to remove, or at least mitigate, the atmosphere from ground-based CMB experiments.

We present a field-level inference framework that reconstructs the static CMB field and per-scan frozen atmospheric screens from time-ordered data. As a proof of concept, we work under the simple assumptions of known wind advection velocity and two-point statistics priors for both the CMB and the atmospheric components and derive the optimal maximum a posteriori (MAP) and the maximum likelihood (ML) estimators for the CMB map. We provide a geometric interpretation of the optimizer and discuss the quality of the reconstruction as a function of the atmosphere amplitude and detector noise. We find that for most Fourier modes, the uncertainty almost exclusively depends only on the detector noise even when the atmosphere amplitude dominates both the CMB and the detector noise by orders of magnitude.

7.2 Forward model

To reconstruct the true CMB signal from the observation, we first need to understand the physical processes and the observational pipeline through which the signal is generated and the data is collected.

7.2.1 Time-ordered data model

We consider a set of time-ordered data (TOD) vectors $\{\mathbf{d}_s\}$ indexed by scan s. The TOD of each scan contains samples indexed by time t and detector α. Under the

frozen-screen approximation, the atmosphere is a rigid screen advected across the field of view by a known transverse wind $\mathbf{v}_s$.

For the temperature-only case, we discretize the sky into $N_{\rm pix}$ pixels. Let $\mathbf{c} \in \mathbb{R}^{N_{\rm pix}}$ denote the static CMB field, and let $\mathbf{a}_s^0 \in \mathbb{R}^{N_{\rm atm}}$ denote the discretized map of atmospheric screen at reference time $t = 0$ for scan s. In general, $N_{\rm atm}$ need not equal $N_{\rm pix}$; for a moving frozen screen it is convenient to define $\mathbf{a}_s^0$ on an extended pixel grid so that advection remains in-domain over the scan duration. The TOD forward model is simply

$$\mathbf{d}_s = P_s\,\mathbf{c} + W_s\,\mathbf{a}_s^0 + \boldsymbol{\epsilon}_s, \qquad \boldsymbol{\epsilon}_s \sim \mathcal{N}(0, N = \sigma_n^2 I). \tag{7.1}$$

Here $P_s \in \mathbb{R}^{N_{\rm TOD} \times N_{\rm pix}}$ is the pointing operator for scan s, and $W_s \in \mathbb{R}^{N_{\rm TOD} \times N_{\rm atm}}$ applies wind advection and samples the drifting atmospheric screen along the scan. The noise vector $\boldsymbol{\epsilon}_s$ models white instrumental noise, and N is its covariance.

We consider a South Pole Telescope-like mock observation campaign. For each scan s, the telescope moves along a constant-elevation azimuthal trajectory with a fixed sampling cadence. The scanning speed and the focal-plane layout determine the pointing operator P_s. The wind velocity $\mathbf{v}_s$ determines the advection operator W_s. We assume both are known perfectly.

In the remainder of the paper, we only discuss the scalar temperature field for simplicity. The extension to polarization is in principle straightforward, obtained by stacking the Stokes components into a single sky vector and promoting P_s and W_s to block operators that encode detector polarization angles and beam response. However, we will leave this study to future work.

7.2.2 Atmospheric emission modeling

We model atmospheric emission and ice-crystal scattering in the upper atmosphere as a statistically isotropic Gaussian random field that is statistically independent across scans. In the chosen pixel basis for $\mathbf{a}_s^0$, we adopt the discrete Gaussian prior

$$\mathbf{a}_s^0 \sim \mathcal{N}(0, C_a). \tag{7.2}$$

We also adopt a Gaussian prior for the CMB field,

$$\mathbf{c} \sim \mathcal{N}(0, C_c). \tag{7.3}$$

Here $C_a \in \mathbb{R}^{N_{\mathrm{atm}} \times N_{\mathrm{atm}}}$ and $C_c \in \mathbb{R}^{N_{\mathrm{pix}} \times N_{\mathrm{pix}}}$ are covariance matrices in the corresponding pixel bases, i.e. $(C_a)_{ij} \equiv \langle a_i a_j \rangle$ and $(C_c)_{ij} \equiv \langle c_i c_j \rangle$ for the discretized fields. Equivalently, C_a and C_c are the pixel-by-pixel correlation matrices implied by the assumed stationary spectra and the chosen discretization (pixelization and beam). We assume both C_a and C_c are known perfectly.

To specify C_a, we use a Kolmogorov angular power spectrum model [326],

$$C_\ell^{TT} = B_{T,\nu}^2 \, g_\nu^2 \, \sin(\epsilon_{\mathrm{el}})^{1-\beta} \left(\frac{\ell}{2\pi} \right)^{-\beta}, \tag{7.4}$$

$$\tag{7.5}$$

where $\beta = 11/3$, $B_{T,\nu}$ is the empirically measured, frequency-dependent power spectrum amplitude, $g_\nu \equiv dT_{\mathrm{CMB}}/dT_{\mathrm{RJ}}$ converts from Rayleigh–Jeans brightness temperature to CMB thermodynamic units, and ϵ_{el} is the elevation of the line of

Drifting Atmosphere

Time	Pixel 1	Pixel 2
$t=0$	$c_1 + a_0$	c_2
$t=1$	c_1	$c_2 + a_0$

Static Atmosphere

Time	Pixel 1	Pixel 2
$t=0$	$c_1 + a_0$	c_2
$t=1$	$c_1 + a_0$	c_2

Table 7.1: The TOD forward model for a two-pixel, two-timestep experiment. We denote the static CMB values as c_1, c_2 and the atmospheric screen at $t = 0$ have a single degree of freedom a_0. **Left:** The atmosphere drifts to the right by one pixel per timestep. **Right:** The atmosphere is static relative to the sky.

sight. At 220 GHz, the empirical median amplitude reported in Coerver *et al.* [326] is $B^2_{T,220} = 43.0\ \mathrm{mK}^2$. We assume the atmosphere has the same power spectrum for all scans for simplicity, although this is not a limitation of the method in general.

7.3 A two-pixel example

Before proposing the estimator in full generality, let us consider a simple two-pixel model that illustrate the intuition of our method. Consider two detectors staring at two sky pixels at two timesteps with no scanning motion. Each TOD sample has independent Gaussian noise of variance σ_n^2. Let $\mathbf{c} = [c_1, c_2]^\top$ denote the static CMB pixel values, and let the atmospheric screen at $t = 0$ have a single degree of freedom a_0 that drifts rightward one pixel per timestep. The TOD vector $\mathbf{d} = [d_{01}, d_{02}, d_{11}, d_{12}]^\top$ (subscripts denote time, pixel) follows the linear model

$$\mathbf{d} = P\mathbf{c} + Wa_0 + \boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma_n^2 I) \tag{7.6}$$

with

$$P = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad W = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}. \tag{7.7}$$

A pictorial representation of the TOD model is shown in the left panel of Tab. 7.1. The pixel values we want to solve for are (c_1, c_2, a_0), which follow the Gaussian likelihood,

$$\log p(\mathbf{d} \mid \mathbf{c}, a_0) = -\frac{1}{2\sigma_n^2} (\mathbf{d} - P\mathbf{c} - Wa_0)^T (\mathbf{d} - P\mathbf{c} - Wa_0) + \text{const.} \tag{7.8}$$

The maximum-likelihood estimator is obtained by solving the normal equations $\frac{\partial}{\partial(\mathbf{c},a_0)} \log p(\mathbf{d} \mid \mathbf{c}, a_0) = 0$, yielding

$$\hat{c}_1 = \frac{1}{4}(d_{01} + 3d_{11} + d_{02} - d_{12}), \tag{7.9}$$

$$\hat{c}_2 = \frac{1}{4}(d_{12} + 3d_{02} + d_{11} - d_{01}), \tag{7.10}$$

$$\hat{a}_0 = \frac{1}{2}(d_{01} + d_{12} - d_{11} - d_{02}). \tag{7.11}$$

The Fisher information matrix for (c_1, c_2, a_0) is

$$F = \frac{1}{\sigma_n^2} \begin{pmatrix} 2 & 0 & 1 \\ 0 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}, \qquad F^{-1} = \frac{\sigma_n^2}{4} \begin{pmatrix} 3 & 1 & -2 \\ 1 & 3 & -2 \\ -2 & -2 & 4 \end{pmatrix}. \tag{7.12}$$

Marginalizing over a_0, the uncertainty on the CMB pixel value is

$$\sigma(c_1) = \sigma(c_2) = \frac{\sqrt{3}}{2}\sigma_n. \tag{7.13}$$

Crucially, this uncertainty depends only on the detector noise σ_n, independent of the atmospheric amplitude a_0. The dynamical distinction between the static sky signal and the moving foreground allows for this separation at the level of the time-ordered data.

Conversely, if the atmosphere is static (right panel of Tab. 7.1), the Fisher matrix becomes singular:

$$F = \frac{1}{\sigma_n^2}\begin{pmatrix} 2 & 0 & 2 \\ 0 & 2 & 0 \\ 2 & 0 & 2 \end{pmatrix}. \tag{7.14}$$

Here, c_1 and a_0 are perfectly degenerate. This model demonstrates that wind advection is the specific mechanism that breaks the degeneracy between the CMB and the atmospheric foreground.

7.4 Joint regression for CMB and per-scan frozen atmosphere

Now we generalize the aforementioned intuition into an estimator of the CMB map with full generality, accounting for scan patterns, multiple scan synthesis, and also the prior on the statistical distribution of both the CMB and the atmosphere maps. The derivation comes in two parts. First, we write down the joint posterior distribution

over the CMB and the atmospheric degrees of freedom. Then, we analytically derive the optimal (MAP and ML) estimators for the CMB signal.

7.4.1 Global optimal CMB reconstruction

We index scans by $s = 1, \ldots, N_{\text{scan}}$. For each scan, the discretized TOD model is

$$\mathbf{d}_s = P\,\mathbf{c} + W_s\,\mathbf{a}_s^0 + \boldsymbol{\epsilon}_s. \tag{7.15}$$

Here $\mathbf{c} \in \mathbb{R}^{N_{\text{pix}}}$ is the static CMB map and $\mathbf{a}_s^0 \in \mathbb{R}^{N_{\text{atm}}}$ is the atmospheric screen at reference time $t = 0$. We assume the scanning pattern is the same for each scan, and hence suppress the dependence of the pointing operator P for conciseness. We assume white instrumental noise $\boldsymbol{\epsilon}_s \sim \mathcal{N}(0, N = \sigma_n^2 I)$, and adopt Gaussian priors $\mathbf{c} \sim \mathcal{N}(0, C_c)$ and $\mathbf{a}_s^0 \sim \mathcal{N}(0, C_a)$.

Since atmospheres are independent across scans, the joint posterior factorizes as

$$p(\mathbf{c}, \{\mathbf{a}_s^0\}|\{\mathbf{d}_s\}) \propto p(\mathbf{c}) \prod_{s=1}^{N_{\text{scan}}} p(\mathbf{d}_s|\mathbf{c}, \mathbf{a}_s^0) p(\mathbf{a}_s^0). \tag{7.16}$$

The log-posterior is

$$\begin{aligned}
&\log p(\mathbf{c}, \{\mathbf{a}_s^0\} \mid \{\mathbf{d}_s\}) \\
&\qquad = -\frac{1}{2}\mathbf{c}^T C_c^{-1}\mathbf{c} \\
&\qquad\quad - \frac{1}{2}\sum_{s=1}^{N_{\rm scan}} \left[(\mathbf{d}_s - P\mathbf{c} - W_s\mathbf{a}_s^0)^T N^{-1}(\mathbf{d}_s - P\mathbf{c} - W_s\mathbf{a}_s^0) + (\mathbf{a}_s^0)^T C_a^{-1}\mathbf{a}_s^0\right] \\
&\qquad\quad + \text{const.}
\end{aligned} \tag{7.17}$$

The unknown parameters in the joint posterior are $\mathbf{c}$ and $\{\mathbf{a}_s^0\}$ and live in a $(N_{\rm pix} + N_{\rm scan}N_{\rm atm})$ dimensional space, making joint estimation computationally formidable. However, since we only seek the CMB map, we can marginalize over the atmospheric degrees of freedom. This is especially convenient because atmospheres are independent across scans, allowing per-scan marginalization. The marginal posterior over CMB is

$$p(\mathbf{c}|\{\mathbf{d}_s\}) \propto p(\mathbf{c}) \prod_{s=1}^{N_{\rm scan}} \left[\int d\mathbf{a}_s^0\, p(\mathbf{d}_s|\mathbf{c}, \mathbf{a}_s^0) p(\mathbf{a}_s^0)\right]. \tag{7.18}$$

Defining the effective covariance

$$\tilde{N}_s \equiv N + W_s C_a W_s^T, \tag{7.19}$$

the marginal log-posterior is (see Appendix 7.7.2 for derivation)

$$\log p(\mathbf{c}|\{\mathbf{d}_s\}) = -\frac{1}{2}\mathbf{c}^T C_c^{-1}\mathbf{c} - \frac{1}{2}\sum_{s=1}^{N_{\rm scan}} (\mathbf{d}_s - P\mathbf{c})^T \tilde{N}_s^{-1}(\mathbf{d}_s - P\mathbf{c}) + \text{const.} \tag{7.20}$$

Physically, the marginal distribution says that effective noise on the CMB pixels consists of white instrumental noise and spatially correlated atmospheric foreground. As we will see shortly, $\tilde{N}_s$ beautifully encodes how well we can recover the different modes of the CMB map, as a function of the wind velocity. The marginal posterior is now only $N_{\rm pix}$ dimensional and can be solved analytically. The MAP estimator is at the critical point of the posterior surface,

$$\frac{\partial}{\partial \mathbf{c}} \log p(\mathbf{c}|\{\mathbf{d}_s\}) = -C_c^{-1}\mathbf{c} + \sum_{s=1}^{N_{\rm scan}} P^T \tilde{N}_s^{-1}(\mathbf{d}_s - P\mathbf{c}) = 0, \tag{7.21}$$

which yields

$$\hat{\mathbf{c}} = \mathrm{Cov}(\hat{\mathbf{c}}) \sum_{s=1}^{N_{\rm scan}} P^T \tilde{N}_s^{-1}\mathbf{d}_s, \quad \mathrm{Cov}(\hat{\mathbf{c}}) = \left(\sum_{s=1}^{N_{\rm scan}} P^T \tilde{N}_s^{-1} P + C_c^{-1} \right)^{-1}. \tag{7.22}$$

In practice, solving Eq. 7.22 as written can be expensive and unstable. We discuss a more numerically advantageous re-formulation in Appendix 7.7.1.

7.4.2 Per-scan CMB reconstruction and optimal synthesis

We now show that, the optimal solutions in Eq. 7.22 is equivalent to, first reconstructing a low SNR CMB map using each scan independently, and then synthesize the independent reconstructions. This equivalent formulation provides a convenient avenue for us to understand how the data from each scan adds information to the final CMB reconstruction.

In this approach, we must take care not to double count the two-point priors

on the CMB field. For each scan s, define the effective posterior distribution that includes a fraction $1/N_{\rm scan}$ of the CMB prior,

$$\mathcal{L}_s(\mathbf{c}, \mathbf{a}_s^0) = -\frac{1}{2}(\mathbf{d}_s - P\mathbf{c} - W_s\mathbf{a}_s^0)^T N^{-1}(\mathbf{d}_s - P\mathbf{c} - W_s\mathbf{a}_s^0) - \frac{1}{2}(\mathbf{a}_s^0)^T C_a^{-1}\mathbf{a}_s^0 - \frac{1}{2N_{\rm scan}}\mathbf{c}^T C_c^{-1}\mathbf{c}. \tag{7.23}$$

Marginalizing the atmosphere $\mathbf{a}_s^0$ as in Appendix 7.7.2, the per-scan MAP estimator and its posterior covariance are,

$$\hat{\mathbf{c}}_s = \mathrm{Cov}(\hat{\mathbf{c}}_s)\, P^T \tilde{N}_s^{-1}\mathbf{d}_s, \qquad \mathrm{Cov}(\hat{\mathbf{c}}_s) \equiv \left(P^T \tilde{N}_s^{-1} P + \frac{1}{N_{\rm scan}} C_c^{-1}\right)^{-1}. \tag{7.24}$$

To optimally synthesize the information from all scans, we combine the per-scan estimates using inverse-variance weighting,

$$\hat{\mathbf{c}} = \left(\sum_{s=1}^{N_{\rm scan}} \mathrm{Cov}(\hat{\mathbf{c}}_s)^{-1}\right)^{-1} \sum_{s=1}^{N_{\rm scan}} \mathrm{Cov}(\hat{\mathbf{c}}_s)^{-1}\hat{\mathbf{c}}_s, \qquad \mathrm{Cov}(\hat{\mathbf{c}}) = \left(\sum_{s=1}^{N_{\rm scan}} \mathrm{Cov}(\hat{\mathbf{c}}_s)^{-1}\right)^{-1}, \tag{7.25}$$

which is algebraically identical to the global MAP solution (again, see Appendix 7.7.3 for derivation).

7.5 The geometry and properties of the optimal estimators

We now show that the optimal estimators have intuitive geometric meanings and behaviors in different noise limits. For clarity, let the sky and the detector array be one dimensional. Each detector is static and observing a fixed sky pixel. The frozen screen atmosphere translates to the right at constant velocity. We consider a single

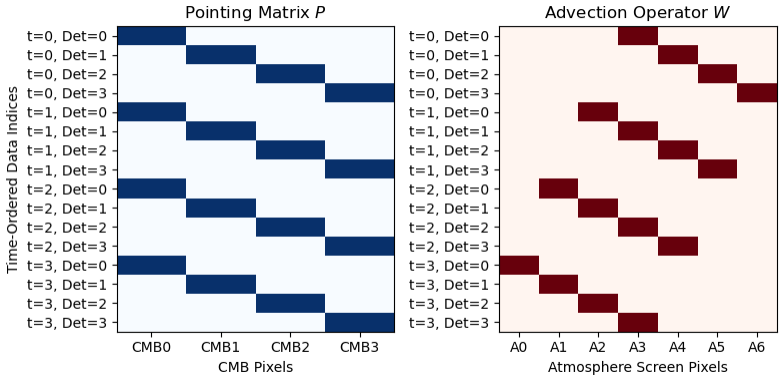

Figure 7.1: An example of the pointing and advection operators P and W, which map the static CMB signal and the frozen atmosphere screen at $t = 0$ to the TOD, respectively. Rows index TOD samples (t, p) $(t = 0, 1, 2, 3)$ and columns index CMB pixels (for P, left panel) or atmosphere pixels at $t = 0$ (for W, right panel); nonzero entries indicate which pixel contributes to each TOD sample. Note that, since the atmospheric screen is moving, the atmosphere map we need to model (7 pixels) extends beyond the spatial domain of the CMB pixels that we observe (4 pixels).

scan and suppress the scan index henceforth. We show an example of the binary matrices P and W in Fig. 7.1.

We will now analyze the limiting cases where the detector noise or the atmopsheric emission dominates. The important equations are, for MAP, $\hat{\mathbf{c}} = \mathrm{Cov}(\hat{\mathbf{c}})\, P^T \tilde{N}^{-1} \mathbf{d}$, $\mathrm{Cov}(\hat{\mathbf{c}}) = \left(P^T \tilde{N}^{-1} P + C_c^{-1} \right)^{-1}$, and $\tilde{N} = N + W C_a W^T$. The ML solution simply corresponds to a flat prior on $\mathbf{c}$, taking the limit $C_c^{-1} \to 0$.

7.5.1 The large detector noise limit

When the detector noise is large, $\tilde{N} \to \sigma_n^2 I$. The ML estimator becomes

$$\hat{\mathbf{c}} = (P^T P)^{-1} P^T \mathbf{d}. \tag{7.26}$$

This is the standard co-addition map-making formula, where we bin the data into pixels and average them. The uncertainty on the reconstructed CMB pixel is $\mathrm{Cov}(\hat{\mathbf{c}}) = \sigma_n^2/N_{\mathrm{obs}}$, where N_{obs} is the number of time each CMB pixel is observed. The MAP estimator, on the other hand, converges to zero because

$$\hat{\mathbf{c}} = \left(P^T N^{-1} P + C_c^{-1}\right)^{-1} P^T N^{-1} \mathbf{d} \tag{7.27}$$

$$= \left(\frac{1}{\sigma_n^2} P^T P + C_c^{-1}\right)^{-1} \frac{1}{\sigma_n^2} P^T \mathbf{d} \xrightarrow{\sigma_n^2 \to \infty} 0, \tag{7.28}$$

and the covariance $\mathrm{Cov}(\hat{\mathbf{c}}) = C_c$ is completely dominated by the prior.

7.5.2 The large atmosphere emission amplitude limit

The large atmosphere is the more realistic limit and provides a clean geometric picture behind the reconstruction algorithm. We will only discuss the ML estimator, first in real space and then in the Fourier space. The MAP case follow similarly.

Real space

Expanding $\tilde{N}^{-1}$ via the Woodbury identity, we have,

$$\tilde{N}^{-1} = N^{-1} - N^{-1} W \left(C_a^{-1} + W^T N^{-1} W\right)^{-1} W^T N^{-1}. \tag{7.29}$$

If we further assume white detector noise $N = \sigma_n^2 I$, then in the large-atmosphere limit,

$$\lim_{\sigma_a \to \infty} \tilde{N}^{-1} = \frac{1}{\sigma_n^2} \left[I - W (W^T W)^{+} W^T\right], \tag{7.30}$$

where $(\cdot)^+$ denotes the Moore–Penrose pseudoinverse. Remarkably, $\tilde{N}^{-1}$ has the form of a scaled projection operator. More precisely, $\Pi_{\text{atm}} = W(W^TW)^+W^T$ projects onto the column space of W. What is the column space of W? W is a matrix of size $N_{\text{TOD}} \times N_{\text{atm}}$. The column space of W represents all N_{TOD}-dimensional time-stream data vectors that could possibly be created by a frozen atmosphere drifting across the sky (the set of $\mathbf{d}$ that can be written as $W\mathbf{a}_0$ for some $\mathbf{a}_0$). This space only depends on the wind velocity (through the advection operator W) and does *not* depend on the specific realization of the atmosphere.

Consequently, the inverse noise covariance

$$\tilde{N}^{-1} = \frac{1}{\sigma_n^2}\Pi_{\text{atm}}^{\perp} = \frac{1}{\sigma_n^2}\left(I - \Pi_{\text{atm}}\right). \tag{7.31}$$

projects the data vector $\mathbf{d}$ onto its orthogonal complement of the column space. Simply put, $\tilde{N}^{-1}$ keeps the parts of the data that is impossible to be generated from a drifting atmosphere. With this in mind, the ML estimator,

$$\hat{\mathbf{c}}_{\text{ML}} = \underbrace{\sigma_n^2(P^T\Pi_{\text{atm}}^{\perp}P)^+}_{\text{Covariance}}\underbrace{\sigma_n^{-2}P^T\Pi_{\text{atm}}^{\perp}\mathbf{d}}_{\text{Projected Data}}, \tag{7.32}$$

projects out all modes in the time-stream data that can be generated by the frozen atmosphere screen.

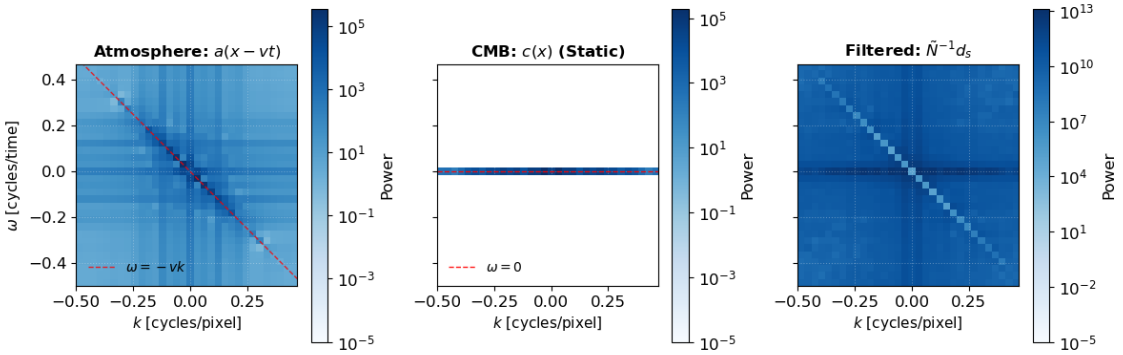

Figure 7.2: Fourier-space decomposition of the one-dimensional idealized model's TOD. The horizontal axis is spatial frequency k (cycles/pixel) and the vertical axis is temporal frequency ω (cycles/sample); color indicates squared Fourier amplitude (arbitrary units). From left to right, the panels show the atmospheric contribution, the static CMB contribution, and the inverse-covariance filtered TOD $\tilde{N}^{-1}\mathbf{d}$. The drifting atmosphere is confined to the dispersion relation $\omega = -vk$ while the static CMB lies on $\omega = 0$. Applying the optimal filter $\tilde{N}^{-1}$ suppresses power along the atmospheric line but leaves most of the CMB signal invariant.

Fourier space

The Fourier space representation of the TOD provides another geometric perspective. Consider the continuous limit of the regularly spaced, static detector array. As such, we can use the detector index and the physical coordinate x interchangeably, and view the TOD as a function of position x and time t.

First, we visualize the power distribution of the atmospheric and the CMB signals in Fourier space. The atmospheric contribution is a frozen profile $a^0(x)$ advected with velocity v, given by $a(x,t) = a^0(x - vt)$. Its spacetime Fourier transform is

$$\tilde{a}(k,\omega) = \int dt\, e^{-i\omega t} \int dx\, e^{-ikx} a^0(x - vt) = \tilde{a}^0(k)\,(2\pi)\delta(\omega + kv). \tag{7.33}$$

In other words, the atmosphere is strictly confined to the line $\omega = -kv$ in the (k, ω) plane, as shown in the left panel of Fig. 7.2.

Conversely, the static CMB signal $c(x, t) = c(x)$ is constant in time. A similar derivation shows that its spectral power is confined to the line $\omega = 0$ (middle panel of Fig. 7.2),

$$\tilde{c}(k, \omega) = \tilde{c}(k)\,(2\pi)\delta(\omega). \tag{7.34}$$

In this idealized continuous limit, the projection operator $\Pi_{\rm atm}$ projects the TOD onto the column space of the advection operator W. Since the range of W consists of all valid atmosphere-like time-streams, and all such time-streams are non-zero only on the dispersion relation $\omega = -kv$, $\Pi_{\rm atm}$ is diagonal in the Fourier basis and acts as a binary mask,

$$\tilde{\Pi}_{\rm atm}(k, \omega) = \begin{cases} 1 & \text{if } \omega = -kv \\ \\ 0 & \text{otherwise} \end{cases}. \tag{7.35}$$

The inverse covariance filter, which is proportional to our estimator operator, $\sigma_n^{-2}\Pi_{\rm atm}^{\perp} = \sigma_n^{-2}(I - \Pi_{\rm atm})$ is the complement of this mask. It zeros out the Fourier coefficients on the atmospheric line $\omega = -kv$ and preserves data everywhere else. This is reflected in the right panel of Fig. 7.2.

Now, we are ready to derive the uncertainty on the CMB pixel by computing the eigenvalues of the Fisher information matrix

$$F = \sigma_n^{-2} P^T \Pi_{\rm atm}^{\perp} P = \frac{N_{\rm obs}}{\sigma_n^2} \Pi_{\rm atm}^{\perp}, \tag{7.36}$$

where we used the commutativity[2] of P and $\Pi^{\perp}_{\rm atm}$ and that $P^T P = N_{\rm obs} I$ for a uniform scan. Therefore, the eigenvalue of the Fisher matrix, λ_k, is:

$$\lambda_k = \frac{N_{\rm obs}}{\sigma_n^2} \times \begin{cases} 0 & \text{if } kv = 0 \\ 1 & \text{if } kv \neq 0 \end{cases}. \tag{7.37}$$

Intuitively, the covariance of the reconstructed CMB signal is diagonal in Fourier space, and each wavevector k has its own uncertainty. At the monopole ($k = 0$) or when there is no wind ($v = 0$), we have no information on the CMB from the data – since the atmosphere overlaps with the CMB at $\omega = 0$ and the filter blocks this frequency, these modes are prior-dominated in MAP or have infinite uncertainty in ML. In contrast, when $kv \neq 0$, the atmosphere can be perfectly removed down to the detector noise level (variance $N_{\rm obs}/\sigma_n^2$), independent of the atmosphere strength.

Figure. 7.3 explicitly verifies this theoretical observation through a simple numerical experiment. Here, the atmosphere advects rightwards, and we took the TOD over 20 time steps with no scanning. We set the CMB RMS σ_c to be 100 times the detector noise, and perform reconstruction at different atmospheric RMS amplitudes from $\sigma_a = 10^{-2}\sigma_c$ to $\sigma_a = 10^2\sigma_c$. The left panel of Fig. 7.3 shows the posterior variances of the different modes (i.e., eigenvalues of the posterior covariance). The right panel shows the spatial amplitude variation of the eigenmodes themselves. The eigenmodes are approximately sine waves. Low frequency modes (near the monopole) have reconstruction uncertainties scaling with the severity of the atmospheric contamination

2. Since they are both diagonal in Fourier space. Equivalently, they are translationally invariant linear operators.

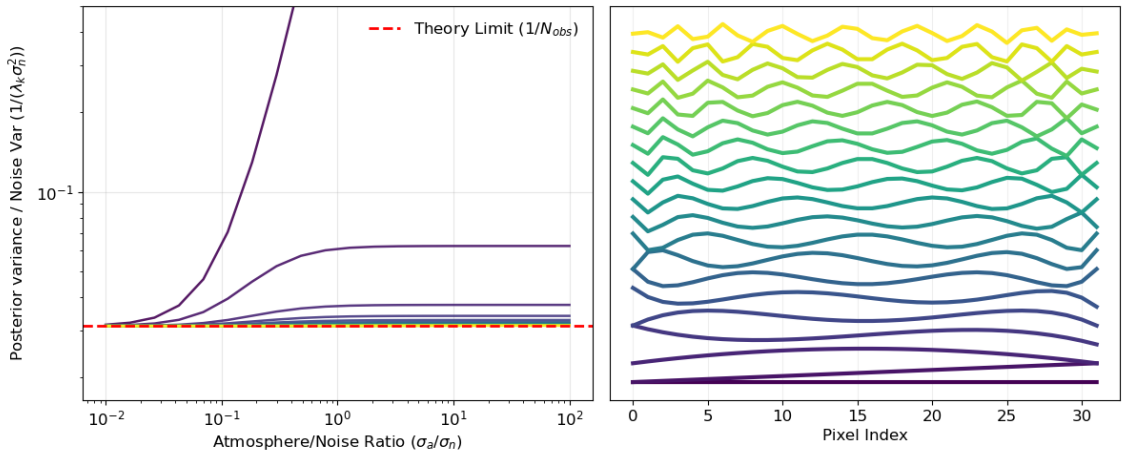

Figure 7.3: Uncertainty scaling and eigenmodes of the posterior covariance in the 1D model (top 20 most uncertain modes). The left panel shows the normalized posterior variances, i.e. $1/(\lambda_k\sigma_n^2)$, versus atmosphere-to-noise amplitude ratio. The right panel shows the corresponding eigenmodes, with the same color scale, ordered from most uncertain (bottom) to most certain (top).

because they overlap with the degenerate subspace, and are prior-dominated in MAP or formally unconstrained in ML. The higher frequency modes have reconstruction uncertainties almost constant as the atmospheric contamination varies across orders of magnitude. These modes are orthogonal to the atmospheric column space and have uncertainty that only depends on the detector noise, $1/\lambda_k \simeq \sigma_n^2/N_{\rm obs}$. Note that the numerical experiment suffers from discretization. In the continuum limit, the eigenmodes converge to sine waves, and the uncertainty becomes independent of the atmospheric foreground except for the monopole. This is the main result of this work.

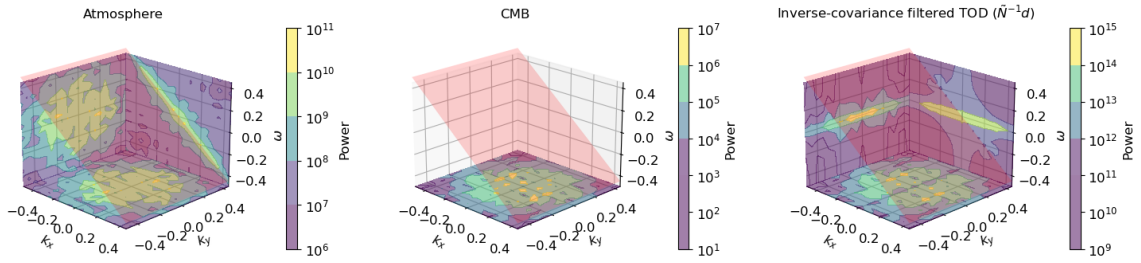

Figure 7.4: Three-dimensional Fourier power decomposition of the idealized 2D TOD example, where the wind is along the $+x$ direction. Axes are spatial frequencies (k_x, k_y) (cycles/pixel) and temporal frequency ω (cycles/sample), with color indicating squared Fourier amplitude (arbitrary units). From left to right, the first panel shows the atmospheric contribution, which lies on the plane $\omega = -\mathbf{k} \cdot \mathbf{v}$ for a frozen screen advected at velocity $\mathbf{v}$ in the $+x$ direction (indicated by the translucent red plane). The middle panel shows the static sky contribution on $\omega = 0$. The right panel shows the inverse-covariance filtered TOD $\tilde{N}^{-1}\mathbf{d}$. The wall projections show the Fourier power integrated over the omitted dimension. Filtering suppresses power near the atmospheric manifold, leaving power concentrated near $\omega = 0$ except along the degeneracy line $\mathbf{k} \cdot \mathbf{v} = 0$.

7.6 Two dimensional model and multi-scan synthesis

In two spatial dimensions, we can clearly see how the multi-scan synthesis (which is formally equivalent to the global optimal solution) combines information from different scans. Again we consider detectors regularly spaced on a two-dimensional grid with no scanning motion.

The Fourier-space representation of the TOD is then three-dimensional, indexed by (k_x, k_y, ω). An example where the atmosphere advects in the $+x$ direction is shown in Fig. 7.4. For a frozen screen advected at constant velocity $\mathbf{v}$, the atmospheric TOD obeys the dispersion relation $\omega = -\mathbf{k} \cdot \mathbf{v}$ in the (k_x, k_y, ω) space (left panel, also shown as the red plane), while a static sky remains confined to $\omega = 0$ (middle panel). The action of the inverse-covariance operator $\tilde{N}^{-1}$ (right panel) is to

filter out all signals that lie in the atmospheric subspace (the red plane). However, unlike the one-dimensional case, the CMB subspace and the advecting atmosphere no longer intersect only at the monopole, but along the line $\mathbf{k} \cdot \mathbf{v} = 0$. Information along this line is suppressed by the optimal reconstruction and hence lost. Physically, this line corresponds to spatial modes orthogonal to the wind direction, and these modes are indistinguishable from a static contaminant and are therefore strongly suppressed. Indeed, in Fig. 7.5, we show the 5 most uncertain eigenmodes of the posterior covariance. All these modes vary primarily in the y direction, perpendicular to the wind velocity in $+x$.

However, not all hope is lost to recover the information along $\mathbf{k} \cdot \mathbf{v} = 0$. Since each scan has independent wind velocity realizations, the information removed during the reconstruction of one scan can be compensated in another scan, through inverse variance weighted averaging in Eq. 7.25. Technically, only the monopole remains formally unconstrained during multi-scan synthesis, which is expected since we cannot disentangle the mean atmospheric and CMB emission amplitudes. Fig. 7.6 illustrates this in the idealized 2D model: reconstructing from a scan with wind along $+x$ suppresses modes near $k_x = 0$, while wind along $+y$ suppresses modes near $k_y = 0$. When scans have different wind directions, the corresponding degenerate subspaces differ in (k_x, k_y, ω) space, so combining information across scans recovers modes that were degenerate in any one scan, leaving only the true monopole unconstrained.

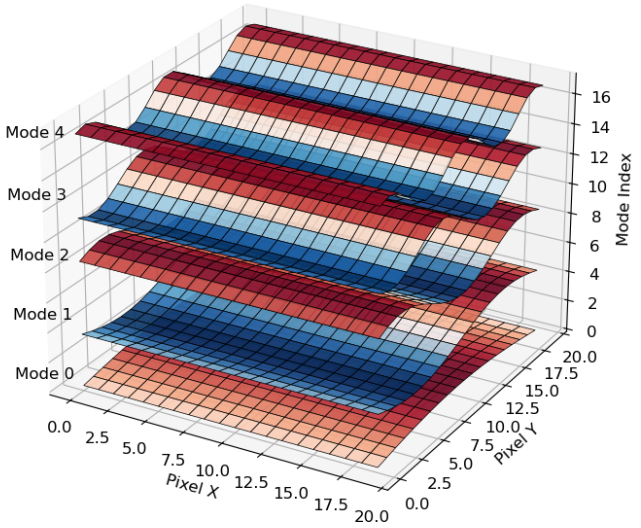

Figure 7.5: Representative high-variance eigenmodes of the 2D posterior covariance for a single scan with a fixed wind in the positive x direction. The horizontal axes are pixel coordinates (x, y). Each surface shows one eigenmode of the posterior covariance (ordered from below, from largest to smaller posterior variance), plotted as a height field over (x, y) with color indicating the signed eigenmode amplitude. The least constrained modes are those whose Fourier representation overlaps the single-scan degeneracy $\mathbf{k} \cdot \mathbf{v} = 0$, which for wind along $+x$ corresponds to stripe-like modes with $k_x \simeq 0$ (variation primarily in y).

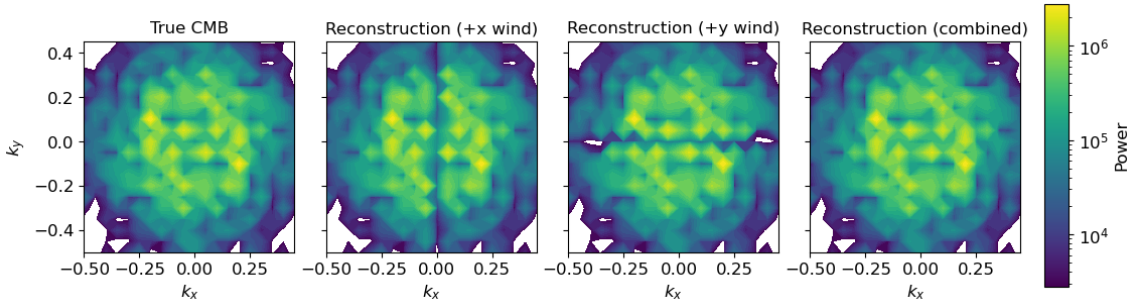

Figure 7.6: Idealized reconstruction comparison for two scans with different wind directions, shown in the $\omega = 0$ plane in (k_x, k_y) space (cycles/pixel) with color indicating squared Fourier amplitude (arbitrary units). From left to right, the panels show the true CMB contribution, the reconstruction from a scan with wind along $+x$, the reconstruction from a scan with wind along $+y$, and the optimally combined reconstruction. Each single scan loses modes near its degeneracy line $\mathbf{k} \cdot \mathbf{v} = 0$ (e.g. $k_x = 0$ for wind along $+x$), while combining scans with different wind directions restores these modes up to the monopole.

7.7 Appendix

7.7.1 Numerical implementation

For conciseness, we only discuss the global MAP solution. ML and per-scan algorithms follow similarly.

Computing the per-scan inverse covariance matrix $P^T \tilde{N}_s^{-1} P$ and data vector $P^T \tilde{N}_s^{-1} \mathbf{d}_s$ requires applying the effective inverse covariance $\tilde{N}_s^{-1}$, where $\tilde{N}_s = N + W_s C_a W_s^T \in \mathbb{R}^{N_{\mathrm{TOD}} \times N_{\mathrm{TOD}}}$. Direct inversion is expensive. Instead, we apply the Woodbury identity,

$$\tilde{N}_s^{-1} = N^{-1} - N^{-1} W_s M^{-1} W_s^T N^{-1}, \qquad M = W_s^T N^{-1} W_s + C_a^{-1}, \tag{7.38}$$

where $M \in \mathbb{R}^{N_{\text{atm}} \times N_{\text{atm}}}$. Since $N = \sigma_n^2 I$ is diagonal, the bottleneck becomes applying C_a^{-1} (which is efficient via FFTs) and solving linear systems with M. We compute $M^{-1}v$ by a Cholesky factorization $M = LL^T$ followed by triangular solves. These per-scan quantities $P^T \tilde{N}_s^{-1} P$ and $P^T \tilde{N}_s^{-1} \mathbf{d}_s$ are accumulated across all scans, and the global MAP solution is obtained by solving $(\sum_s P^T \tilde{N}_s^{-1} P + C_c^{-1})\hat{\mathbf{c}} = \sum_s P^T \tilde{N}_s^{-1} \mathbf{d}_s$.

7.7.2 Marginalization over atmospheric screens

We derive the marginal posterior $p(\mathbf{c}|\{\mathbf{d}_s\})$ by analytically integrating out the atmospheric fields $\{\mathbf{a}_s^0\}$. Since atmospheres are independent across scans, we marginalize each scan separately. For scan s, the relevant terms in the log-posterior are

$$\mathcal{L}_s(\mathbf{c}, \mathbf{a}_s^0) = -\frac{1}{2}\left[(\mathbf{d}_s - P\mathbf{c} - W_s\mathbf{a}_s^0)^T N^{-1}(\mathbf{d}_s - P\mathbf{c} - W_s\mathbf{a}_s^0) + (\mathbf{a}_s^0)^T C_a^{-1}\mathbf{a}_s^0\right]. \tag{7.39}$$

Collecting all terms quadratic in $\mathbf{a}_s^0$:

$$\mathcal{L}_s(\mathbf{c}, \mathbf{a}_s^0) = -\frac{1}{2}\left[(\mathbf{a}_s^0)^T(W_s^T N^{-1} W_s + C_a^{-1})\mathbf{a}_s^0 - 2(\mathbf{d}_s - P\mathbf{c})^T N^{-1} W_s\, \mathbf{a}_s^0\right] + f(\mathbf{c}), \tag{7.40}$$

where $f(\mathbf{c}) = -\frac{1}{2}(\mathbf{d}_s - P\mathbf{c})^T N^{-1}(\mathbf{d}_s - P\mathbf{c})$ is independent of $\mathbf{a}_s^0$. We rewrite $\mathcal{L}_s(\mathbf{c}, \mathbf{a}_s^0)$ in the quadratic form to perform the Gaussian integral

$$\mathcal{L}_s(\mathbf{c}, \mathbf{a}_s^0) = -\frac{1}{2}\left[\left(\mathbf{a}_s^0 - \boldsymbol{\mu}_s\right)^T \Lambda_s \left(\mathbf{a}_s^0 - \boldsymbol{\mu}_s\right) - \boldsymbol{\mu}_s^T \Lambda_s \boldsymbol{\mu}_s\right] + f(\mathbf{c}), \tag{7.41}$$

where we define the inverse covariance matrix $\Lambda_s = W_s^T N^{-1} W_s + C_a^{-1}$ and $\boldsymbol{\mu}_s = \Lambda_s^{-1} W_s^T N^{-1}(\mathbf{d}_s - P\mathbf{c})$ is the conditional mean of $\mathbf{a}_s^0$ given $\mathbf{c}$ and $\mathbf{d}_s$. Integrating over $\mathbf{a}_s^0$ yields

$$\begin{aligned}
\log p(\mathbf{d}_s|\mathbf{c}) &= \frac{1}{2}\boldsymbol{\mu}_s^T \Lambda_s \boldsymbol{\mu}_s + f(\mathbf{c}) + \text{const} \\
&= -\frac{1}{2}(\mathbf{d}_s - P\mathbf{c})^T \left[N^{-1} - N^{-1} W_s \Lambda_s^{-1} W_s^T N^{-1}\right](\mathbf{d}_s - P\mathbf{c}) + \text{const} \qquad (7.42) \\
&= -\frac{1}{2}(\mathbf{d}_s - P\mathbf{c})^T \tilde{N}_s^{-1}(\mathbf{d}_s - P\mathbf{c}) + \text{const}, \qquad (7.43)
\end{aligned}$$

where the last line follows from the Woodbury identity, and the effective covariance is defined as $\tilde{N}_s = N + W_s C_a W_s^T$.

7.7.3 The equivalence between global MAP estimator and the synthesized per-scan solutions

We show the optimally combined per-scan solutions is equivalent to the global MAP. For the combined estimator, the posterior covariance after inverse-variance weighting

is

$$\mathrm{Cov}(\hat{\mathbf{c}}) = \left(\sum_{s=1}^{N_{\mathrm{scan}}} \mathrm{Cov}(\hat{\mathbf{c}}_s)^{-1} \right)^{-1} \tag{7.44}$$

$$= \left(\sum_{s=1}^{N_{\mathrm{scan}}} \left(P^T \tilde{N}_s^{-1} P + \frac{1}{N_{\mathrm{scan}}} C_c^{-1} \right) \right)^{-1} \tag{7.45}$$

$$= \left(\sum_{s=1}^{N_{\mathrm{scan}}} P^T \tilde{N}_s^{-1} P + C_c^{-1} \right)^{-1}, \tag{7.46}$$

For the point estimate, note that $\mathrm{Cov}(\hat{\mathbf{c}}_s)^{-1}\hat{\mathbf{c}}_s = \mathrm{Cov}(\hat{\mathbf{c}}_s)^{-1}\mathrm{Cov}(\hat{\mathbf{c}}_s)P^T\tilde{N}_s^{-1}\mathbf{d}_s = P^T\tilde{N}_s^{-1}\mathbf{d}_s$, therefore

$$\hat{\mathbf{c}} = \left(\sum_{s=1}^{N_{\mathrm{scan}}} \mathrm{Cov}(\hat{\mathbf{c}}_s)^{-1} \right)^{-1} \sum_{s=1}^{N_{\mathrm{scan}}} \mathrm{Cov}(\hat{\mathbf{c}}_s)^{-1}\hat{\mathbf{c}}_s \tag{7.47}$$

$$= \left(\sum_{s=1}^{N_{\mathrm{scan}}} P^T \tilde{N}_s^{-1} P + C_c^{-1} \right)^{-1} \sum_{s=1}^{N_{\mathrm{scan}}} P^T \tilde{N}_s^{-1} \mathbf{d}_s, \tag{7.48}$$

which also matches the global MAP point estimate.

REFERENCES

[1] Markus Michael Rau, Roohi Dalal, *et al.*, Monthly Notices of the Royal Astronomical Society **524**, 5109 (2023).

[2] Euclid Collaboration, Mischa Knabenhans, *et al.*, Monthly Notices of the Royal Astronomical Society **484**, 5509 (2019).

[3] Dhayaa Anbajagane, Shivam Pandey, *et al.*, Map-level baryonification: Efficient modelling of higher-order correlations in the weak lensing and thermal Sunyaev-Zeldovich fields (2024).

[4] Shivam Pandey, Jaime Salcido, *et al.*, arXiv e-prints , arXiv:2401.18072 (2024).

[5] Scott Dodelson, *Modern cosmology* (Academic Press, San Diego, Calif, 2003).

[6] Scott Dodelson, *Gravitational Lensing* (2017).

[7] Peter Schneider, in *33rd Advanced Saas Fee Course on Gravitational Lensing: Strong, Weak, and Micro* (2006) pp. 269–451, arXiv:astro-ph/0509252 .

[8] Planck Collaboration, Y. Akrami, *et al.*, Astronomy & Astrophysics **641**, A4 (2020).

[9] G. Hinshaw, D. Larson, *et al.*, The Astrophysical Journal Supplement Series **208**, 19 (2013).

[10] DES Collaboration, T. M. C. Abbott, *et al.*, Physical Review D **105**, 023520 (2022).

[11] E. Camphuis, W. Quan, *et al.*, Spt-3g d1: Cmb temperature and polarization power spectra and cosmology from 2019 and 2020 observations of the spt-3g main field (2025), arXiv:2506.20707 [astro-ph.CO] .

[12] Hironao Miyatake, Sunao Sugiyama, *et al.*, Hyper Suprime-Cam Year 3 Results: Cosmology from Galaxy Clustering and Weak Lensing with HSC and SDSS using the Emulator Based Halo Model (2023).

[13] Sunao Sugiyama, Hironao Miyatake, *et al.*, Physical Review D **108**, 123521 (2023).

[14] Planck Collaboration, N. Aghanim, *et al.*, Astronomy & Astrophysics **641**, A6 (2020).

[15] M. Gatti, G. Campailla, *et al.*, Dark energy survey year 3 results: simulation-based cosmological inference with wavelet harmonics, scattering transforms, and moments of weak lensing mass maps ii. cosmological results (2025).

[16] Antony Lewis and Sarah Bridle, Physical Review D **66**, 103511 (2002), arXiv:astro-ph/0205436 .

[17] Michael Betancourt, A Conceptual Introduction to Hamiltonian Monte Carlo (2018).

[18] Alan Junzhe Zhou, Xiangchong Li, *et al.*, Physical Review D **110**, 023539 (2024).

[19] Alan Junzhe Zhou, Marco Gatti, *et al.*, Journal of Cosmology and Astroparticle Physics **09**, 073, arXiv:2505.07949 [astro-ph.CO] .

[20] Jens Jasche and Benjamin D. Wandelt, Monthly Notices of the Royal Astronomical Society **432**, 894 (2013).

[21] Jens Jasche and Guilhem Lavaux, Astronomy & Astrophysics **625**, A64 (2019).

[22] M. Millea, C. M. Daley, *et al.*, The Astrophysical Journal **922**, 259 (2021).

[23] Pier Fiedorowicz, Eduardo Rozo, *et al.*, Monthly Notices of the Royal Astronomical Society **512**, 73 (2022).

[24] Alan Junzhe Zhou and Scott Dodelson, Physical Review D **108**, 083506 (2023).

[25] Alan Junzhe Zhou, Scott Dodelson, *et al.*, Phys. Rev. Lett. **135**, 261002 (2025), arXiv:2506.14878 [astro-ph.CO] .

[26] Alan Junzhe Zhou, Yin Li, *et al.*, Journal of Cosmology and Astroparticle Physics **2024** (10), 069.

[27] Radford M. Neal, *MCMC using Hamiltonian dynamics* (Chapman and Hall/CRC, 2011).

[28] Planck Collaboration, N. Aghanim, *et al.*, Astronomy & Astrophysics **641**, A1 (2020).

[29] Planck Collaboration, N. Aghanim, *et al.*, Astronomy & Astrophysics **641**, A8 (2020).

[30] Simone Aiola, Erminia Calabrese, *et al.*, Journal of Cosmology and Astroparticle Physics **2020** (12), 047.

[31] Omar Darwish, Mathew S Madhavacheril, *et al.*, Monthly Notices of the Royal Astronomical Society **500**, 2250 (2020).

[32] F. Bianchini, W. L. K. Wu, *et al.*, The Astrophysical Journal **888**, 119 (2020).

[33] L. Balkenhol, D. Dutcher, *et al.*, A measurement of the CMB temperature power spectrum and constraints on cosmology from the SPT-3G 2018 TT/TE/EE data set (2022).

[34] Catherine Heymans, Tilman Tröster, *et al.*, Astronomy & Astrophysics **646**, A140 (2021).

[35] Takashi Hamana, Masato Shirasaki, *et al.*, Publications of the Astronomical Society of Japan **72**, 10.1093/pasj/psz138 (2020).

[36] T. M. C. Abbott, M. Aguena, *et al.*, Joint analysis of DES year 3 data and CMB lensing from SPT and Planck III: Combined cosmological constraints (2022).

[37] Max Tegmark, The Astrophysical Journal **480**, L87 (1997).

[38] Benjamin D. Wandelt, David L. Larson, *et al.*, Physical Review D **70**, 10.1103/physrevd.70.083511 (2004).

[39] D. L. Larson, H. K. Eriksen, *et al.*, The Astrophysical Journal **656**, 653 (2007).

[40] H. K. Eriksen, J. B. Jewell, *et al.*, The Astrophysical Journal **676**, 10 (2008).

[41] D. L. Larson, H. K. Eriksen, *et al.*, The Astrophysical Journal **656**, 653 (2007).

[42] Ethan Anderes, Benjamin D. Wandelt, *et al.*, The Astrophysical Journal **808**, 152 (2015).

[43] Planck Collaboration, N. Aghanim, *et al.*, Astronomy & Astrophysics **641**, A5 (2020).

[44] Eleni Tsaprazi, Nhat-Minh Nguyen, *et al.*, Journal of Cosmology and Astroparticle Physics **08** (08), 003.

[45] A. Loureiro, L. Whiteway, *et al.*, Almanac: Weak Lensing power spectra and map inference on the masked sphere (2022).

[46] Adam Andrews, Jens Jasche, *et al.*, Bayesian field-level inference of primordial non-gaussianity using next-generation galaxy surveys (2022).

[47] Pier Fiedorowicz, Eduardo Rozo, *et al.*, arXiv e-prints , arXiv:2210.12280 (2022).

[48] Jens Jasche and Guilhem Lavaux, Monthly Notices of the Royal Astronomical Society **447**, 1204 (2015).

[49] Doogesh Kodi Ramanah, Guilhem Lavaux, *et al.*, Astronomy & Astrophysics **621**, A69 (2019).

[50] Wayne Hu, The Astrophysical Journal Letters **557**, L79 (2001).

[51] Marius Millea, Ethan Anderes, *et al.*, Physical Review D **100**, 023509 (2019).

[52] Marius Millea, Ethan Anderes, *et al.*, Physical Review D **102**, 123542 (2020).

[53] Qianjun Hang, Shadab Alam, *et al.*, Monthly Notices of the Royal Astronomical Society **501**, 1481 (2021).

[54] K. M. Gorski, E. Hivon, *et al.*, The Astrophysical Journal **622**, 759 (2005).

[55] Andrea Zonca, Leo Singer, *et al.*, Journal of Open Source Software **4**, 1298 (2019).

[56] Radford Neal, in *Handbook of Markov Chain Monte Carlo* (Chapman and Hall/CRC, 2011) pp. 113–162.

[57] Matthew D. Hoffman and Andrew Gelman, arXiv e-prints , arXiv:1111.4246 (2011).

[58] Eli Bingham, Jonathan P. Chen, *et al.*, Pyro: Deep Universal Probabilistic Programming (2018).

[59] Du Phan, Neeraj Pradhan, *et al.*, Composable effects for flexible and accelerated probabilistic programming in numpyro (2019).

[60] Planck Collaboration, Y. Akrami, *et al.*, Planck 2018 results. ix. constraints on primordial non-gaussianity (2019).

[61] George B. Rybicki and William H. Press, The Astrophysical Journal **398**, 169 (1992).

[62] Andrija Kostic, Nhat-Minh Nguyen, *et al.*, Consistency tests of field level inference with the EFT likelihood (2022).

[63] L.A. Kofman and A.A. Starobinskii, Soviet Astronomy Letters **11**, 271 (1985).

[64] Robert G. Crittenden and Neil Turok, Physical Review Letters **76**, 575 (1996).

[65] Pablo Fosalba and Enrique Gaztanaga, Monthly Notices of the Royal Astronomical Society **350**, L37 (2004).

[66] Stephen Boughn and Robert Crittenden, Nature **427**, 45 (2004).

[67] R. Scranton, A. J. Connolly, *et al.*, Physical evidence for dark energy (2003).

[68] Niayesh Afshordi, Physical Review D **70**, 10.1103/physrevd.70.083536 (2004).

[69] A. Manzotti and S. Dodelson, Physical Review D **90**, 123009 (2014).

[70] R. Cawthon, J. Elvin-Poole, *et al.*, Monthly Notices of the Royal Astronomical Society **513**, 5517 (2022).

[71] G. Giannini, A. Alarcon, *et al.*, arXiv e-prints , arXiv:2209.05853 (2022).

[72] Supranta S Boruah, Tim Eifler, *et al.*, Monthly Notices of the Royal Astronomical Society **518**, 4818 (2022).

[73] M Rodrí guez-Monroy, N Weaverdyck, *et al.*, Monthly Notices of the Royal Astronomical Society **511**, 2665 (2022).

[74] Jean-Eric Campagne, François Lanusse, *et al.*, Jax-cosmo: An end-to-end differentiable and gpu accelerated cosmology library (2023).

[75] Daniel J. Eisenstein and Wayne Hu, The Astrophysical Journal **511**, 5 (1999).

[76] Supranta Sarma Boruah, Eduardo Rozo, *et al.*, Map-based cosmology inference with lognormal cosmic shear maps (2022).

[77] Takahiro Nishimichi, Masahiro Takada, *et al.*, The Astrophysical Journal **884**, 29 (2019).

[78] Justin Alsing, Alan Heavens, *et al.*, Monthly Notices of the Royal Astronomical Society **466**, 3272 (2016).

[79] Adrian E. Bayer, Chirag Modi, *et al.*, Joint velocity and density reconstruction of the universe with nonlinear differentiable forward modeling (2022).

[80] Chirag Modi, François Lanusse, *et al.*, Cosmicrim : Reconstructing early universe by combining differentiable simulations with recurrent inference machines (2021).

[81] Jens Jasche, Florent Leclercq, *et al.*, Journal of Cosmology and Astroparticle Physics **2015** (01), 036.

[82] Guilhem Lavaux and Jens Jasche, Monthly Notices of the Royal Astronomical Society **455**, 3169 (2016).

[83] Julia Stadler, Fabian Schmidt, *et al.*, Journal of Cosmology and Astroparticle Physics **2023** (10), 069.

[84] Fabian Schmidt, Journal of Cosmology and Astroparticle Physics **2021** (04), 033.

[85] Fabian Schmidt, Franz Elsner, *et al.*, Journal of Cosmology and Astroparticle Physics **2019** (01), 042.

[86] Franz Elsner, Fabian Schmidt, *et al.*, Journal of Cosmology and Astroparticle Physics **2020** (01), 029.

[87] Justin Alsing, Alan F. Heavens, *et al.*, Monthly Notices of the Royal Astronomical Society **466**, 3272 (2017).

[88] Natalia Porqueres, Alan Heavens, *et al.*, Field-level inference of cosmic shear with intrinsic alignments and baryons (2023).

[89] Supranta S. Boruah and Eduardo Rozo, Map-based cosmology inference with weak lensing – information content and its dependence on the parameter space (2023).

[90] Florent Leclercq and Alan Heavens, Monthly Notices of the Royal Astronomical Society **506**, L85 (2021).

[91] Nhat-Minh Nguyen, Fabian Schmidt, *et al.*, Journal of Cosmology and Astroparticle Physics **2021** (03), 058.

[92] Hiroaki Aihara, Nobuo Arimoto, *et al.*, Publications of the Astronomical Society of Japan **70**, S4 (2018).

[93] Xiangchong Li, Hironao Miyatake, *et al.*, Publications of the Astronomical Society of Japan **74**, 421 (2022).

[94] D. Spergel, N. Gehrels, *et al.*, arXiv e-prints (2015).

[95] Xiangchong Li, Tianqing Zhang, *et al.*, Hyper Suprime-Cam Year 3 Results: Cosmology from Cosmic Shear Two-point Correlation Functions (2023).

[96] Roohi Dalal, Xiangchong Li, *et al.*, Hyper Suprime-Cam Year 3 Results: Cosmology from Cosmic Shear Power Spectra (2023).

[97] Matthias Bartelmann and Peter Schneider, Physics Reports **340**, 291 (2001).

[98] D. Nelson Limber, The Astrophysical Journal **117**, 134 (1953).

[99] Antony Lewis, Anthony Challinor, *et al.*, The Astrophysical Journal **538**, 473 (2000).

[100] Jia Liu, Andrea Petri, *et al.*, Physical Review D **91**, 063507 (2015).

[101] Zack Li, Jia Liu, *et al.*, Physical Review D **99**, 063527 (2019).

[102] Vanessa Böhm, Stefan Hilbert, *et al.*, Physical Review D **96**, 123510 (2017).

[103] S. Hilbert, J. Hartlap, *et al.*, Astronomy & Astrophysics **536**, A85 (2011).

[104] Henrique S. Xavier, Filipe B. Abdalla, *et al.*, Monthly Notices of the Royal Astronomical Society **459**, 3693 (2016).

[105] Rachel Mandelbaum, Barnaby Rowe, *et al.*, The Astrophysical Journal Supplement Series **212**, 5 (2014).

[106] M. A. Troxel and Mustapha Ishak, Physics Reports **558**, 1 (2015).

[107] E. Krause, X. Fang, *et al.*, arXiv e-prints , arXiv:2105.13548 (2021).

[108] Ryuichi Takahashi, Takashi Hamana, *et al.*, The Astrophysical Journal **850**, 24 (2017).

[109] Masato Shirasaki, Takashi Hamana, *et al.*, Monthly Notices of the Royal Astronomical Society **486**, 52 (2019).

[110] Andrea Zonca, Leo Singer, *et al.*, Journal of Open Source Software **4**, 1298 (2019).

[111] The LSST Dark Energy Science Collaboration, Rachel Mandelbaum, *et al.*, arXiv e-prints (2018).

[112] Matthew D. Hoffman and Andrew Gelman, Journal of Machine Learning Research **15**, 1593 (2014).

[113] Nick Kaiser and Gordon Squires, The Astrophysical Journal **404**, 441 (1993).

[114] James W. Kirchner, Physical Review E **71**, 066110 (2005).

[115] Emiliano Sefusatti, Martin Crocce, *et al.*, Monthly Notices of the Royal Astronomical Society **460**, 3624 (2016).

[116] Nicolas Tessore, Arthur Loureiro, *et al.*, The Open Journal of Astrophysics **6**, 10.21105/astro.2302.01942 (2023).

[117] Takahiko Matsubara, The Astrophysical Journal Supplement Series **170**, 1 (2007).

[118] Jian Qin, Jun Pan, *et al.*, Monthly Notices of the Royal Astronomical Society **514**, 1548 (2022).

[119] Tianqing Zhang, Xiangchong Li, *et al.*, arXiv e-prints , arXiv:2212.03257 (2022).

[120] Takahiro Nishimichi, Masahiro Takada, *et al.*, The Astrophysical Journal **884**, 29 (2019).

[121] Yosuke Kobayashi, Takahiro Nishimichi, *et al.*, Physical Review D **105**, 083517 (2022).

[122] A. Kiessling, A. F. Heavens, *et al.*, Monthly Notices of the Royal Astronomical Society **414**, 2235 (2011).

[123] P. Fosalba, E. Gaztanaga, *et al.*, Monthly Notices of the Royal Astronomical Society **447**, 1319 (2015).

[124] R. J. Sgier, A. Réfrégier, *et al.*, Journal of Cosmology and Astroparticle Physics **2019** (01), 044.

[125] Kyle Cranmer, Johann Brehmer, *et al.*, Proceedings of the National Academy of Sciences **117**, 30055 (2020).

[126] Yin Li, Chirag Modi, *et al.*, Differentiable Cosmological Simulation with Adjoint Method (2022).

[127] Yin Li, Libin Lu, *et al.*, pmwd: A Differentiable Cosmological Particle-Mesh N-body Library (2022).

[128] Huiyuan Wang, H. J. Mo, *et al.*, The Astrophysical Journal **794**, 94 (2014).

[129] Uroš Seljak, Grigor Aslanyan, *et al.*, Journal of Cosmology and Astroparticle Physics **2017** (12), 009.

[130] Fabian Schmidt, Franz Elsner, *et al.*, Journal of Cosmology and Astroparticle Physics **2019** (01), 042.

[131] Francisco-Shu Kitaura, Metin Ata, *et al.*, Monthly Notices of the Royal Astronomical Society **502**, 3456 (2021).

[132] Metin Ata, Francisco-Shu Kitaura, *et al.*, Monthly Notices of the Royal Astronomical Society **500**, 3194 (2021).

[133] Denise Lanzieri, François Lanusse, *et al.*, Forecasting the power of Higher Order Weak Lensing Statistics with automatically differentiable simulations (2023).

[134] Nhat-Minh Nguyen, Fabian Schmidt, *et al.*, arXiv preprint arXiv:2403.03220 (2024).

[135] S. Dodelson, E. W. Kolb, *et al.*, Physical Review D **72**, 103004 (2005).

[136] Asantha Cooray and Wayne Hu, The Astrophysical Journal **574**, 19 (2002).

[137] E. Krause and C. M. Hirata, Astronomy & Astrophysics **523**, A28 (2010).

[138] Geraint Pratten and Antony Lewis, Journal of Cosmology and Astroparticle Physics **2016** (08), 047.

[139] Giulio Fabbian, Matteo Calabrese, *et al.*, Journal of Cosmology and Astroparticle Physics **2018** (2), 050.

[140] Andrea Petri, Zoltán Haiman, *et al.*, Physical Review D **95**, 123503 (2017).

[141] Giulio Fabbian, Antony Lewis, *et al.*, Journal of Cosmology and Astroparticle Physics **2019** (10), 057.

[142] Chihway Chang and Bhuvnesh Jain, Monthly Notices of the Royal Astronomical Society **443**, 102 (2014).

[143] Vanessa Böhm, Yu Feng, *et al.*, MADLens, a python package for fast and differentiable non-Gaussian lensing simulations (2020).

[144] Scott Dodelson, Charles Shapiro, *et al.*, Physical Review D **73**, 023009 (2006).

[145] Bhuvnesh Jain, Uroš Seljak, *et al.*, The Astrophysical Journal **530**, 547 (2000).

[146] Christopher M. Hirata and Uros Seljak, Physical Review D **68**, 083002 (2003).

[147] Stefan Hilbert, Alexandre Barreira, *et al.*, Monthly Notices of the Royal Astronomical Society **493**, 305 (2020).

[148] Dominic Beck, Giulio Fabbian, *et al.*, Physical Review D **98**, 043512 (2018).

[149] Vanessa Böhm, Blake D. Sherwin, *et al.*, Physical Review D **98**, 123510 (2018).

[150] Vanessa Böhm, Chirag Modi, *et al.*, Journal of Cosmology and Astroparticle Physics **2020** (3), 045.

[151] Chris Vale and Martin White, The Astrophysical Journal **592**, 699 (2003).

[152] S. Hilbert, S. D. M. White, *et al.*, Monthly Notices of the Royal Astronomical Society **382**, 121 (2007).

[153] S. Hilbert, J. Hartlap, *et al.*, Astronomy & Astrophysics **499**, 31 (2009).

[154] Masanori Sato, Takashi Hamana, *et al.*, The Astrophysical Journal **701**, 945 (2009).

[155] Matthew R. Becker, Monthly Notices of the Royal Astronomical Society **435**, 115 (2013).

[156] Andrea Petri, Astronomy and Computing **17**, 73 (2016).

[157] Ken Osato, Jia Liu, *et al.*, Monthly Notices of the Royal Astronomical Society **502**, 5593 (2021).

[158] Andrea Petri, Zoltán Haiman, *et al.*, Physical Review D **95**, 123503 (2017).

[159] Chengliang Wei, Guoliang Li, *et al.*, The Astrophysical Journal **853**, 25 (2018).

[160] Kun Xu and Yipeng Jing, The Astrophysical Journal **915**, 75 (2021).

[161] J. Jiménez-Vicente and E. Mediavilla, The Astrophysical Journal **941**, 80 (2022).

[162] Xingpao Suo, Xi Kang, *et al.*, The Astrophysical Journal **948**, 56 (2023).

[163] H. M. P. Couchman, Andrew J. Barber, *et al.*, Monthly Notices of the Royal Astronomical Society **308**, 180 (1999).

[164] Baojiu Li, Lindsay J. King, *et al.*, Monthly Notices of the Royal Astronomical Society **415**, 881 (2011).

[165] Alexandre Barreira, Claudio Llinares, *et al.*, Journal of Cosmology and Astroparticle Physics **2016** (05), 001.

[166] M. Killedar, P. D. Lasky, *et al.*, Monthly Notices of the Royal Astronomical Society **420**, 155 (2012).

[167] Rennan Bar-Kana, *Gravitational lensing as a probe of dark matter, the distance scale, and gravitational waves in the universe*, Ph.D. thesis, Massachusetts Institute of Technology (1997).

[168] Thomas Quinn, Neal Katz, *et al.*, Time stepping N-body simulations (1997).

[169] Prasenjit Saha and Scott Tremaine, The Astronomical Journal **104**, 1633 (1992).

[170] Volker Springel, Rüdiger Pakmor, *et al.*, Monthly Notices of the Royal Astronomical Society **506**, 2871 (2021).

[171] Haruo Yoshida, Celestial Mechanics and Dynamical Astronomy **56**, 27 (1993).

[172] Peter Schneider, Jürgen Ehlers, *et al.*, *Gravitational Lenses*, edited by I. Appenzeller, G. Börner, *et al.*, Astronomy and Astrophysics Library (Springer Berlin Heidelberg, Berlin, Heidelberg, 1992).

[173] R. W. Hockney and J. W. Eastwood, *Computer Simulation Using Particles* (CRC Press, 2021).

[174] Reuben Epstein and Irwin I. Shapiro, Physical Review D **22**, 2947 (1980).

[175] Planck Collaboration, P. A. R. Ade, *et al.*, Astronomy & Astrophysics **594**, A13 (2016).

[176] Jill P. Naiman, Annalisa Pillepich, *et al.*, Monthly Notices of the Royal Astronomical Society **477**, 1206 (2018).

[177] Volker Springel, Rüdiger Pakmor, *et al.*, Monthly Notices of the Royal Astronomical Society **475**, 676 (2018).

[178] Federico Marinacci, Mark Vogelsberger, *et al.*, Monthly Notices of the Royal Astronomical Society 10.1093/mnras/sty2206 (2018).

[179] Dylan Nelson, Annalisa Pillepich, *et al.*, Monthly Notices of the Royal Astronomical Society **475**, 624 (2018).

[180] Annalisa Pillepich, Dylan Nelson, *et al.*, Monthly Notices of the Royal Astronomical Society **475**, 648 (2018).

[181] Andrea Petri, Zoltán Haiman, *et al.*, Physical Review D **93**, 063524 (2016).

[182] Yucheng Zhang, Yin Li, *et al.*, Neural and symbolic optimization of cosmological particle-mesh simulation (in prep).

[183] Ryuichi Takahashi, Masanori Sato, *et al.*, The Astrophysical Journal **761**, 152 (2012).

[184] Lingyi Hu, Alan Heavens, *et al.*, Journal of Cosmology and Astroparticle Physics **2022** (02), 009.

[185] P. Schneider, Astronomy & Astrophysics **143**, 413 (1985).

[186] Roger Blandford and Ramesh Narayan, The Astrophysical Journal **310**, 568 (1986).

[187] A. Amon, D. Gruen, *et al.*, Physical Review D **105**, 023514 (2022).

[188] L. F. Secco, S. Samuroff, *et al.*, Physical Review D **105**, 023515 (2022).

[189] Dark Energy Survey and Kilo-Degree Survey Collaboration, T. M. C. Abbott, *et al.*, The Open Journal of Astrophysics **6**, 36 (2023).

[190] D. Anbajagane, C. Chang, *et al.*, arXiv e-prints , arXiv:2502.17677 (2025).

[191] Angus H. Wright, Benjamin Stolzner, *et al.*, arXiv e-prints , arXiv:2503.19441 (2025).

[192] Biwei Dai and Uros Seljak, Multiscale Flow for Robust and Optimal Cosmological Analysis (2024).

[193] Janis Fluri, Tomasz Kacprzak, *et al.*, Physical Review D **100**, 063514 (2019).

[194] M. Gatti, B. Jain, *et al.*, Physical Review D **106**, 083509 (2022).

[195] N Jeffrey, L Whiteway, *et al.*, Monthly Notices of the Royal Astronomical Society **536**, 1303 (2024).

[196] Gabriela A. Marques, Jia Liu, *et al.*, Monthly Notices of the Royal Astronomical Society **528**, 4513 (2024).

[197] Camila P. Novaes, Leander Thiele, *et al.*, arXiv e-prints , arXiv:2409.01301 (2024).

[198] Sihao Cheng, Gabriela A. Marques, *et al.*, Journal of Cosmology and Astroparticle Physics **2025** (1), 006.

[199] Dhayaa Anbajagane, Chihway Chang, *et al.*, Journal of Cosmology and Astroparticle Physics **2024** (3), 062.

[200] Xiangkun Liu, Baojiu Li, *et al.*, Physical Review Letters **117**, 051101 (2016).

[201] Christopher T. Davies, Joachim Harnois-Déraps, *et al.*, Monthly Notices of the Royal Astronomical Society **533**, 3546 (2024).

[202] Oleg Y. Gnedin, Andrey V. Kravtsov, *et al.*, The Astrophysical Journal **616**, 16 (2004).

[203] Alan R. Duffy, Joop Schaye, *et al.*, Monthly Notices of the Royal Astronomical Society **405**, 2161 (2010).

[204] Marcel P. Van Daalen, Joop Schaye, *et al.*, Monthly Notices of the Royal Astronomical Society **415**, 3649 (2011).

[205] E. Semboloni, H. Hoekstra, *et al.*, Monthly Notices of the Royal Astronomical Society **434**, 148 (2013).

[206] Dhayaa Anbajagane, August E. Evrard, *et al.*, Monthly Notices of the Royal Astronomical Society **509**, 3441 (2022).

[207] Mufan Shao, Dhayaa Anbajagane, *et al.*, arXiv e-prints , arXiv:2212.05964 (2022).

[208] Ian G McCarthy, Jaime Salcido, *et al.*, Monthly Notices of the Royal Astronomical Society **526**, 5494 (2023).

[209] Jeger C Broxterman, Matthieu Schaller, *et al.*, Monthly Notices of the Royal Astronomical Society **529**, 2309 (2024).

[210] Matthieu Schaller, Joop Schaye, *et al.*, Monthly Notices of the Royal Astronomical Society **539**, 1337 (2025).

[211] Matthieu Schaller and Joop Schaye, An analytic redshift-independent formulation of baryonic effects on the matter power spectrum (2025).

[212] Elisabetta Semboloni, Henk Hoekstra, *et al.*, Monthly Notices of the Royal Astronomical Society **417**, 2020 (2011).

[213] Mijin Yoon, M. James Jee, *et al.*, The Astrophysical Journal **870**, 111 (2019).

[214] Alexandra Amon and George Efstathiou, Monthly Notices of the Royal Astronomical Society **516**, 5355 (2022).

[215] D. Anbajagane, C. Chang, *et al.*, Monthly Notices of the Royal Astronomical Society **526**, 5530 (2023).

[216] Giulia Piccirilli, Matteo Zennaro, *et al.*, arXiv e-prints , arXiv:2502.17339 (2025).

[217] Ottavia Truttero, Joe Zuntz, *et al.*, The Open Journal of Astrophysics **8**, 19 (2025).

[218] Mark Vogelsberger, Federico Marinacci, *et al.*, Nature Reviews Physics **2**, 42 (2020).

[219] Marco Velliscig, Marcel P. van Daalen, *et al.*, Monthly Notices of the Royal Astronomical Society **442**, 2641 (2014).

[220] Marcel P. Van Daalen and Joop Schaye, Monthly Notices of the Royal Astronomical Society **452**, 2247 (2015).

[221] Stijn N B Debackere, Joop Schaye, *et al.*, Monthly Notices of the Royal Astronomical Society **492**, 2285 (2020).

[222] Dhayaa Anbajagane, August E. Evrard, *et al.*, Monthly Notices of the Royal Astronomical Society **495**, 686 (2020).

[223] S. H. Lim, D. Barnes, *et al.*, Monthly Notices of the Royal Astronomical Society **504**, 5131 (2021).

[224] Elizabeth Lee, Dhayaa Anbajagane, *et al.*, Monthly Notices of the Royal Astronomical Society **517**, 5303 (2022).

[225] Richard Stiskalek, Deaglan J. Bartlett, *et al.*, Monthly Notices of the Royal Astronomical Society **514**, 4026 (2022).

[226] Dhayaa Anbajagane, Han Aung, *et al.*, Monthly Notices of the Royal Astronomical Society **510**, 2980 (2022).

[227] Matthew Gebhardt, Daniel Anglés-Alcázar, *et al.*, arXiv e-prints , arXiv:2307.11832 (2023).

[228] N. E. Chisari, M. L. A. Richardson, *et al.*, Monthly Notices of the Royal Astronomical Society **480**, 3962 (2018).

[229] Marcel P van∼Daalen, Ian G McCarthy, *et al.*, Monthly Notices of the Royal Astronomical Society **491**, 2424 (2020).

[230] Francisco Villaescusa-Navarro, Daniel Anglés-Alcázar, *et al.*, The Astrophysical Journal **915**, 71 (2021).

[231] A. J. Mead, T. Tröster, *et al.*, Astronomy & Astrophysics **641**, A130 (2020).

[232] Ian G. McCarthy, Joop Schaye, *et al.*, Monthly Notices of the Royal Astronomical Society **465**, 2936 (2017).

[233] Aurel Schneider and Romain Teyssier, Journal of Cosmology and Astroparticle Physics **2015** (12), 049.

[234] Aurel Schneider, Romain Teyssier, *et al.*, Journal of Cosmology and Astroparticle Physics **2019** (3), 020.

[235] Giovanni Aricò, Raul E Angulo, *et al.*, Monthly Notices of the Royal Astronomical Society **495**, 4800 (2020).

[236] Sambit K. Giri and Aurel Schneider, Journal of Cosmology and Astroparticle Physics **2021** (12), 046.

[237] A. Chen, G. Aricò, *et al.*, Monthly Notices of the Royal Astronomical Society **518**, 5340 (2023).

[238] Giovanni Aricò, Raul E. Angulo, *et al.*, DES Y3 cosmic shear down to small scales: constraints on cosmology and baryons (2023).

[239] L. Bigwood, A. Amon, *et al.*, arXiv e-prints , arXiv:2404.06098 (2024).

[240] Giovanni Aricò, Raul E. Angulo, *et al.*, Monthly Notices of the Royal Astronomical Society **503**, 3596 (2021).

[241] Max E. Lee, Tianhuan Lu, *et al.*, Monthly Notices of the Royal Astronomical Society **519**, 573 (2023).

[242] Joop Schaye, Roi Kugel, *et al.*, Monthly Notices of the Royal Astronomical Society **526**, 4978 (2023).

[243] Roi Kugel, Joop Schaye, *et al.*, Monthly Notices of the Royal Astronomical Society **526**, 6103 (2023).

[244] Sihao Cheng, Yuan-Sen Ting, *et al.*, Monthly Notices of the Royal Astronomical Society **499**, 5902 (2020).

[245] E. Allys, T. Marchand, *et al.*, Physical Review D **102**, 103506 (2020).

[246] Julio F. Navarro, Carlos S. Frenk, *et al.*, The Astrophysical Journal **462**, 563 (1996).

[247] Julio F. Navarro, Carlos S. Frenk, *et al.*, The Astrophysical Journal **490**, 493 (1997).

[248] Masamune Oguri and Masahiro Takada, Physical Review D **83**, 023008 (2011).

[249] Benedikt Diemer and Andrey V. Kravtsov, The Astrophysical Journal **799**, 108 (2015).

[250] Chun-Hao To, Shivam Pandey, *et al.*, arXiv e-prints , arXiv:2402.00110 (2024).

[251] Janis Fluri, Tomasz Kacprzak, *et al.*, Physical Review D **105**, 083518 (2022).

[252] Josh Borrow, Matthieu Schaller, *et al.*, Monthly Notices of the Royal Astronomical Society **511**, 2367 (2022).

[253] Matthieu Schaller, Josh Borrow, *et al.*, Monthly Notices of the Royal Astronomical Society **530**, 2378 (2024).

[254] Sylvia Ploeckinger and Joop Schaye, Monthly Notices of the Royal Astronomical Society **497**, 4857 (2020).

[255] Joop Schaye and Claudio Dalla Vecchia, Monthly Notices of the Royal Astronomical Society **383**, 1210 (2008).

[256] Robert P. C. Wiersma, Joop Schaye, *et al.*, Monthly Notices of the Royal Astronomical Society **399**, 574 (2009).

[257] Willem Elbers, Carlos S. Frenk, *et al.*, Monthly Notices of the Royal Astronomical Society **507**, 2614 (2021).

[258] Claudio Dalla Vecchia and Joop Schaye, Monthly Notices of the Royal Astronomical Society **387**, 1431 (2008).

[259] Evgenii Chaikin, Joop Schaye, *et al.*, Monthly Notices of the Royal Astronomical Society **523**, 3709 (2023).

[260] C. M. Booth and Joop Schaye, in *The Monster's Fiery Breath: Feedback in Galaxies, Groups, and Clusters*, American Institute of Physics Conference Series, Vol. 1201, edited by Sebastian Heinz and Eric Wilcots (AIP, 2009) pp. 21–24.

[261] Filip Huško, Cedric G. Lacey, *et al.*, Monthly Notices of the Royal Astronomical Society **516**, 3750 (2022).

[262] A Amon, N C Robertson, *et al.*, Monthly Notices of the Royal Astronomical Society **518**, 477 (2022).

[263] M. Gatti, G. Giannini, *et al.*, arXiv e-prints , arXiv:2012.08569 (2020).

[264] J. Myles, A. Alarcon, *et al.*, Monthly Notices of the Royal Astronomical Society **505**, 4249 (2021).

[265] Jeger C. Broxterman, Matthieu Schaller, *et al.*, Monthly Notices of the Royal Astronomical Society **538**, 755 (2025).

[266] Ananth Tenneti, Nickolay Y. Gnedin, *et al.*, The Astrophysical Journal **834**, 169 (2017).

[267] Adam Soussana, Nora Elisa Chisari, *et al.*, Monthly Notices of the Royal Astronomical Society **492**, 4268 (2020).

[268] M. Gatti, N. Jeffrey, *et al.*, Physical Review D **109**, 063534 (2024).

[269] Stéphane Mallat, Communications on Pure and Applied Mathematics **65**, 1331 (2012).

[270] Julien Carron, The Astrophysical Journal **738**, 86 (2011).

[271] E. Allys, F. Levrier, *et al.*, Astronomy & Astrophysics **629**, A115 (2019).

[272] Stéphane Mallat, Sixin Zhang, *et al.*, Phase Harmonic Correlations and Convolutional Neural Networks (2019).

[273] F. Bernardeau, S. Colombi, *et al.*, Physics Reports **367**, 1 (2002).

[274] Hung-Jin Huang, Tim Eifler, *et al.*, Monthly Notices of the Royal Astronomical Society **488**, 1652 (2019).

[275] Daniela Grandón, Gabriela A. Marques, *et al.*, Impact of baryonic feedback on HSC Y1 weak lensing non-Gaussian statistics (2024).

[276] James Bradbury, Roy Frostig, *et al.*, JAX: Composable transformations of Python+NumPy programs (2018).

[277] Willem Elbers, Carlos S Frenk, *et al.*, Monthly Notices of the Royal Astronomical Society **537**, 2160 (2025).

[278] D. Anbajagane, A. Alarcon, *et al.*, arXiv e-prints , arXiv:2502.17675 (2025).

[279] Angus H. Wright, Hendrik Hildebrandt, *et al.*, arXiv e-prints , arXiv:2503.19440 (2025).

[280] Bhuvnesh Jain and Ludovic Van Waerbeke, The Astrophysical Journal **530**, L1 (2000).

[281] Bing Xu, Naiyan Wang, *et al.*, Empirical Evaluation of Rectified Activations in Convolutional Network (2015).

[282] Nitish Srivastava, Geoffrey Hinton, *et al.*, Journal of Machine Learning Research **15**, 1929 (2014).

[283] Diederik P. Kingma and Jimmy Ba, Adam: A Method for Stochastic Optimization (2017).

[284] G. Hinshaw, D. N. Spergel, *et al.*, The Astrophysical Journal Supplement Series **148**, 135 (2003).

[285] D. N. Spergel, L. Verde, *et al.*, The Astrophysical Journal Supplement Series **148**, 175 (2003).

[286] Angélica De Oliveira-Costa, Max Tegmark, *et al.*, Physical Review D **69**, 063516 (2004).

[287] H. K. Eriksen, A. J. Banday, *et al.*, The Astrophysical Journal **622**, 58 (2005).

[288] Kate Land and Joao Magueijo, Physical Review Letters **95**, 071301 (2005).

[289] Craig J. Copi, Dragan Huterer, *et al.*, Advances in Astronomy **2010**, 847541 (2010).

[290] C. L. Bennett, R. S. Hill, *et al.*, The Astrophysical Journal Supplement Series **192**, 17 (2011).

[291] Planck Collaboration, Y. Akrami, *et al.*, Astronomy & Astrophysics **641**, A7 (2020).

[292] DESI Collaboration, A. G. Adame, *et al.*, Journal of Cosmology and Astroparticle Physics **2025** (02), 021.

[293] DESI Collaboration, M. Abdul-Karim, *et al.*, DESI DR2 Results II: Measurements of Baryon Acoustic Oscillations and Cosmological Constraints (2025).

[294] Eleonora Di Valentino, Olga Mena, *et al.*, Classical and Quantum Gravity **38**, 153001 (2021).

[295] Marc Kamionkowski and Adam G. Riess, The Hubble Tension and Early Dark Energy (2022).

[296] Edmund J. Copeland, M. Sami, *et al.*, International Journal of Modern Physics D **15**, 1753 (2006).

[297] Asantha Cooray, Daniel E. Holz, *et al.*, Journal of Cosmology and Astroparticle Physics **2010** (11), 015.

[298] A. Maleknejad, M. M. Sheikh-Jabbari, *et al.*, Physics Reports **528**, 161 (2013).

[299] Jacques Colin, Roya Mohayaee, *et al.*, Monthly Notices of the Royal Astronomical Society **414**, 264 (2011).

[300] Jaiseung Kim and Eiichiro Komatsu, Physical Review D **88**, 101301 (2013).

[301] Daniela Saadeh, Stephen M. Feeney, *et al.*, Physical Review Letters **117**, 131302 (2016).

[302] John Soltis, Arya Farahi, *et al.*, Physical Review Letters **122**, 091301 (2019).

[303] Pablo Fosalba and Enrique Gaztanaga, Monthly Notices of the Royal Astronomical Society **504**, 5840 (2021).

[304] Jessica A. Cowell, Suhail Dhawan, *et al.*, Monthly Notices of the Royal Astronomical Society **526**, 1482 (2023).

[305] S. Yeung and M. C. Chu, Physical Review D **105**, 083508 (2022).

[306] Balakrishna S. Haridasu, Paolo Salucci, *et al.*, Radial Tully-Fisher relation and the local variance of Hubble parameter (2024).

[307] C. Gimeno-Amo, F. K. Hansen, *et al.*, Exploring Statistical Isotropy in Planck Data Release 4: Angular Clustering and Cosmological Parameter Variations Across the Sky (2025).

[308] Andrea Zonca, Leo Singer, *et al.*, Journal of Open Source Software **4**, 1298 (2019).

[309] Adam G. Riess, Wenlong Yuan, *et al.*, The Astrophysical Journal Letters **934**, L7 (2022).

[310] Lam Hui and Patrick B. Greene, Physical Review D **73**, 123526 (2006).

[311] Dragan Huterer, Daniel L. Shafer, *et al.*, Journal of Cosmology and Astroparticle Physics **2017** (05), 015.

[312] Khaled Said, Cullan Howlett, *et al.*, DESI Peculiar Velocity Survey – Fundamental Plane (2024).

[313] Joseph B. Jensen, John P. Blakeslee, *et al.*, The Astrophysical Journal Supplement Series **255**, 21 (2021).

[314] Khaled Said, Matthew Colless, *et al.*, Monthly Notices of the Royal Astronomical Society **497**, 1275 (2020).

[315] David Alonso, Javier Sanchez, *et al.*, Monthly Notices of the Royal Astronomical Society **484**, 4127 (2019).

[316] J. Hartlap, T. Schrabback, *et al.*, Astronomy & Astrophysics **504**, 689 (2009).

[317] Scott Dodelson and Michael D. Schneider, Physical Review D **88**, 063537 (2013).

[318] Christoph Saulder, Cullan Howlett, *et al.*, Monthly Notices of the Royal Astronomical Society **525**, 1106 (2023).

[319] Jian-Ping Hu, Yu-Yan Wang, *et al.*, Astronomy & Astrophysics **681**, A88 (2024).

[320] S. E. Church, Monthly Notices of the Royal Astronomical Society **272**, 551 (1995).

[321] Oliver P. Lay and Nils W. Halverson, The Astrophysical Journal **543**, 787–798 (2000).

[322] G. I. Taylor, Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences **164**, 476 (1938).

[323] R. S. Bussmann, W. L. Holzapfel, *et al.*, The Astrophysical Journal **622**, 1343 (2005).

[324] J. Errard, P. A. R. Ade, *et al.*, The Astrophysical Journal **809**, 63 (2015).

[325] Thomas W. Morris, Ricardo Bustos, *et al.*, Physical Review D **105**, 042004 (2022).

[326] A. Coerver, J. A. Zebrowski, *et al.*, The Astrophysical Journal **982**, 15 (2025).