# Towards Optimal Cosmological Analysis with Simulation-Based Inference

by

Biwei Dai

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Physics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Uroš Seljak, Chair
Professor Joshua S. Bloom
Professor Liang Dai

Summer 2024

Towards Optimal Cosmological Analysis with Simulation-Based Inference

Copyright 2024
by
Biwei Dai

Abstract

Towards Optimal Cosmological Analysis with Simulation-Based Inference

by

Biwei Dai

Doctor of Philosophy in Physics

University of California, Berkeley

Professor Uroš Seljak, Chair

The large-scale structures (LSS) of the Universe contain significant information about the cosmos, providing insights into its composition and the underlying physical laws. Current and upcoming cosmological surveys aim to measure the LSS and the evolution of the universe with multiple probes, allowing us to constrain cosmological parameters to high precision and search for deviations from the standard cosmological model. To realize the full potential of these datasets, it is crucial to develop robust analysis methods capable of extracting the maximum amount of information.

Simulation-based inference (SBI) leverages high-fidelity cosmological simulations for inference and holds promise for extracting rich non-Gaussian information from these data. However, its application is limited by several challenges. This dissertation focuses on addressing these challenges to facilitate the deployment of SBI approaches in upcoming survey analyses.

On the simulation side, running a large number of high-resolution, large-volume cosmological simulations for training the SBI model is computationally challenging. We develop effective machine learning models to improve the modeling of non-linear gravity and baryonic physics in low-resolution fast simulations. By combining these models with fast N-body simulations, we can predict various baryonic observables and accurate weak lensing signals at a low computational cost.

On the inference side, we integrate physical constraints (symmetries) and domain knowledge (the hierarchical structure of the data) into the SBI models, and apply them to learn the field-level data likelihood function for optimal cosmological analysis. Inaccurate modeling of physical processes and systematic effects could bias the SBI constraints. To address this, we use field-level data likelihood and multiscale analysis for anomaly detection of model misspecification, enabling robust SBI analysis. Additionally, we perform a large-scale comparative study to identify the best hyperparameter and loss function choices for optimal SBI performance.

These developments mark a substantial step toward the full deployment of SBI approaches into cosmological survey analysis pipelines, offering the promise of a deeper understanding of our Universe and the potential discovery of new physics beyond the current model.

# Contents

**Contents** **i**

**List of Figures** **iv**

**List of Tables** **ix**

**1 Introduction** **1**
1.1 Large-Scale Structures . . . . . 3
1.2 Cosmological Simulations . . . . . 4
1.3 Cosmological Inference . . . . . 5
1.4 Simulation-Based Inference . . . . . 7
1.5 Dissertation Outline . . . . . 8

**2 A gradient-based method for modeling baryons and matter in halos of fast simulations** **10**
2.1 Introduction . . . . . 11
2.2 Gradient-based learning: theory and motivation . . . . . 12
2.3 Example application: matching a set of hydrodynamical and N-body simulations . 19
2.4 Parameter Selection of PGD . . . . . 24
2.5 Conclusions . . . . . 30
2.6 Appendix A: Simulation data sets . . . . . 31
2.7 Appendix B: Cost Function and Choice Covariance . . . . . 32

**3 High mass and halo resolution from fast low-resolution simulations** **34**
3.1 Introduction . . . . . 34
3.2 Halo statistics and clustering . . . . . 35
3.3 Dark matter statistics . . . . . 45
3.4 Conclusions . . . . . 51
3.5 Appendix A: Halo power spectrum of higher mass thresholds . . . . . 52

**4 MADLens, a python package for fast and differentiable non-Gaussian lensing simulations** **56**
4.1 Introduction . . . . . 56
4.2 Weak Gravitational Lensing, Notation and Conventions . . . . . 58

4.3 MADLens package design . . . 59
4.4 Results . . . 61
4.5 Summary & Outlook . . . 68
4.6 Appendix A: Differentiability with respect to Cosmological Parameters . . . 71
4.7 Appendix B: Details about FastPM and the sub-evolution scheme . . . 74

**5 Learning effective physical laws for generating cosmological hydrodynamics with Lagrangian Deep Learning 77**
5.1 Introduction . . . 78
5.2 Lagrangian Deep Learning . . . 79
5.3 Results . . . 83
5.4 Discussion . . . 91
5.5 Appendix A: Materials and Methods . . . 92

**6 A field-level emulator for modeling baryonic effects across hydrodynamic simulations 97**
6.1 Introduction . . . 97
6.2 Simulations . . . 99
6.3 Baryonic effects at the field-level . . . 104
6.4 Gaussian Process Emulator . . . 105
6.5 Results . . . 108
6.6 Conclusions . . . 114

**7 Translation and Rotation Equivariant Normalizing Flow (TRENF) for Optimal Cosmological Analysis 117**
7.1 Introduction . . . 118
7.2 Method . . . 121
7.3 Results: generative samples in data space and data representation in latent space . . . 125
7.4 Results: likelihood and posterior analysis . . . 128
7.5 Modeling the Data With Mask . . . 132
7.6 Beyond spherical kernels . . . 135
7.7 Discussion . . . 137

**8 Multiscale Flow for Robust and Optimal Cosmological Analysis 140**
8.1 Introduction . . . 140
8.2 Multiresolution Analysis with Fast Wavelet Transform . . . 142
8.3 Multiscale Flow . . . 144
8.4 Results . . . 148
8.5 Discussion . . . 156
8.6 Appendix A: Materials and Methods . . . 157

**9 A comparative study of cosmological constraints from weak lensing using Convolutional Neural Networks 162**

9.1 Introduction . . . . . 163
9.2 Materials and Methods . . . . . 164
9.3 Results . . . . . 170
9.4 Transfer Learning . . . . . 173
9.5 Conclusions . . . . . 177

**10 Conclusions** **179**

**Bibliography** **181**

# List of Figures

1.1 The cosmic history according to the $\Lambda$CDM model from Big Bang to the present day. Credit: NASA. . . . . 2

2.1 The filter as a function of k. . . . . 14
2.2 A visualization of the descent along the gradient of the gravitational potential. . . . . 15
2.3 The power spectra of 10 step FastPM simulation using PGD models with different $\alpha$. . 16
2.4 The projection of the same halo in Illustris, IllustrisDark, and FastPM, before and after applying our models. . . . . 20
2.5 The ratio of matter power spectra (left panel) and cross-correlation coefficients (right panel) of FastPM with 10 steps, 40 steps, and dark-matter-only simulation, before and after using our models, compared to hydrodynamical simulations. . . . . 21
2.6 The halo power spectrum for different halo mass (upper panels), halo cross-correlation coefficient between FastPM and Illustris-Dark (lower left panel) and the ratio of halo mass function between FastPM and Illustris-Dark (lower right panel). . . . . 22
2.7 The spherical averaged halo density profile (upper panels) and enclosed mass as a function of radius (lower panels), in different halo mass bins. . . . . 23
2.8 Visualizations of some large halos and their satellite sub-halos in Illustris and 10 step FastPM. . . . . 25
2.9 Visualizations of some large halos and their satellite sub-halos in Illustris and 40 step FastPM. . . . . 26
2.10 The satellite sub-halo power spectra (upper left), cross-correlation coefficients with the reference simulation (upper right), stochasticity (lower left), and the average numbers per halo (lower right). . . . . 27
2.11 The marginalized posterior probability distribution of the 3 parameters in the potential gradient descent model when fitting the power spectra of 10 steps FastPM against Illustris (blue color) and Illustris-Dark (red color). . . . . 28
2.12 The power spectra of FastPM simulation after applying the PGD model. . . . . 29

3.1 The projected map of the same halo in TNG300-2-Dark (left panel), FastPM (middle panel) and TNG300-3-Dark (right panel). . . . . 38
3.2 The ratio of missed halos as a function of halo particle number in different simulations and various redshifts. . . . . 39

3.3 The ratio of halo bias as a function of abundance measured in different simulations and various redshifts. . . . . 41
3.4 The ratio of halo mass function from different simulations and various redshifts. . . . . 42
3.5 The ratio of halo auto power spectrum from different simulations and various redshifts. 43
3.6 The ratio of halo-matter cross power spectrum from different simulations and various redshifts. . . . . 44
3.7 The cross-correlation coefficient of the reference halo with halos from other simulations in different redshifts. . . . . 45
3.8 The ratio of the halo power spectrum in redshift space from different simulations and various redshifts. . . . . 46
3.9 The matter power spectrum of FastPM simulation, before and after the calibration, and the reference simulation (TNG300 and TNG300-Dark) in different redshifts. . . . . 47
3.10 The redshift distribution of source galaxies of an LSST-like survey (left panel), and the zoomed-in convergence field of FastPM (middle panel) and FastPM with PGD correction (right panel). . . . . 49
3.11 The weak lensing convergence auto power spectrum (upper panel) and cross power spectrum (bottom panel). . . . . 50
3.12 The ratio of halo auto power spectrum in real space from different simulations and various redshifts. . . . . 52
3.13 The same as Figure 3.12, but for halos $M \geq 10^{12} M_{\odot} = 6.8 \times 10^{11} h^{-1} M_{\odot}$. . . . . 53
3.14 The ratio of halo auto power spectrum in redshift space from different simulations and various redshifts, for halos $M \geq 2 \times 10^{11} M_{\odot} = 1.35 \times 10^{11} h^{-1} M_{\odot}$. . . . . 54
3.15 The same as Figure 3.14, but for halos $M \geq 10^{12} M_{\odot} = 6.8 \times 10^{11} h^{-1} M_{\odot}$. . . . . 55

4.1 A MADLens convergence map at $z$=1.0, based on a 3D simulation of side length $256\ \mathrm{Mpc/h}$ and $256^3$ particles. . . . . 62
4.2 PDF of convergence values in the map in Figure 4.1. . . . . 63
4.3 MADLens outputs (box length 512 Mpc/h, $512^3$ particles, with sub-evolution and PGD enhancement) for different source redshifts but same initial conditions (no shot noise subtraction). . . . . 64
4.4 Shot noise and comparison to a theoretical convergence power spectrum based on a HaloFit matter power spectrum: The MADLens simulation traces the HaloFit power spectrum within the accuracy of HaloFit (gray band). . . . . 65
4.5 The PGD enhancement recovers lensing power on small scales in a low resolution simulation. . . . . 66
4.6 Cross correlation between a low resolution MADLens run with and without PGD enhancement, and a simulation with 8 times more particles and no PGD enhancement (same as Fig 4.5). . . . . 67
4.7 The sub-evolution scheme corrects for a systematic overestimation in power that occurs when snapshots are projected at the end of every FastPM step instead of at the exact position/redshift that corresponds to their distance to the observer (light blue). . . . . 68

4.8 The MADLens automatic derivative agrees excellently with the result of finite differencing. . . . 69
4.9 Computation time for a single simulation run with 11 steps and a box size of 256 Mpc/h for different source redshifts. . . . 69
4.10 Scaling of the computation time with the number of processes for a $256^3$ particle/ 256 Mpc/h box simulation. . . . 70
4.11 A comparison of different settings of the simulation package in terms of computation time. . . . 71
4.12 Computation time of vector-Jacobian (vJp) and Jacobian-vector product (Jvp) in units of computation time of the forward model for different source redshifts. . . . 72
4.13 The percent error of the CLASS convergence map power spectrum with respect to the Eisenstein & Hu transfer function convergence maps. . . . 73
4.14 Comparison of the gradient calculation between automatic derivative and finite differencing. . . . 74
4.15 The cross-correlation coefficient of the gradient maps between finite difference and automatic derivative. . . . 75

5.1 Visualization of slices of the simulations. . . . 84
5.2 Comparison of the test set 3D power spectrum (top panel), transfer function (middle panel) and cross correlation coefficient (bottom panel) of the stellar mass overdensity. . . . 86
5.3 Comparison of the test set 2D power spectrum (top panel), transfer function (middle panel) and cross correlation coefficient (bottom panel) of the electron momentum density $n_e v_z$ (proportional to kSZ signal) between the LDL hybrid simulations, TNG300-3 and the target TNG300-1 hydrodynamical simulation. . . . 87
5.4 Comparison of the test set 3D power spectrum (top panel), transfer function (middle panel) and cross correlation coefficient (bottom panel) of the electron pressure $n_e T$ (proportional to tSZ signal) between the LDL hybrid simulations, TNG300-3 and the target TNG300-1 hydrodynamical simulation. . . . 88
5.5 Comparison of the test set 3D power spectrum (top panel), transfer function (middle panel) and cross correlation coefficient (bottom panel) of the gas property $n_e^2 T^{0.5}$ (proportional to X-ray emissivity) between the LDL hybrid simulations, TNG300-3 and the target TNG300-1 hydrodynamical simulation. . . . 89
5.6 The ratio of the test set cross power spectrum of different observables between the LDL hybrid simulations and the target TNG300-1 hydrodynamical simulation. . . . 90

6.1 The matter power spectra ratio observed across suites within CAMELS at various redshifts. . . . 101
6.2 The ratio of matter power spectra in the 200 Astrid test simulations for $z = 0.0$. . . . 102
6.3 Ratios of the matter power spectra in the Astrid $z = 0.0$ CV set. . . . 103
6.4 The cross-correlation coefficients between N-body and hydrodynamic fields across all suites within CAMELS at different redshifts. . . . 106
6.5 Emulator performance on all CAMELS test simulations. . . . 110

6.6 The emulator predictions alongside the reference hydrodynamic simulations for scenarios outside the CAMELS dataset. . . . . 111
6.7 Comparison between our GP emulator and other baryonification emulators - BACCO (left panel), HMcode (middle panel), and BCemu (right panel) - on Astrid test data at $z = 0.0$. . . . . 113
6.8 Comparison of our emulator against BACCO's 7 and 3 parameter versions, HMcode, and BCemu using hydrodynamic simulations outside of CAMELS. . . . . 114
6.9 Field-level improvement comparison in the IllustisTNG simulation suite across different redshifts. . . . . 115

7.1 Test data (left panel) and uncurated TRENF samples (right panel) as a function of conditional variables $\Omega_m$ and $\sigma_8$. . . . . 125
7.2 The pixel probability distribution function (upper left), power spectrum (upper right), and bispectrum (lower panels) of FastPM, TRENF samples, and FastPM data in TRENF latent space. . . . . 127
7.3 Test data (left panel), latent data transformed with correct cosmology ($\Omega_m = 0.309$, $\sigma_8 = 0.816$, middle panel), and latent data transformed with incorrect cosmology ($\Omega_m = 0.4$, $\sigma_8 = 0.5$, right panel). . . . . 128
7.4 The comparison of log-likelihood (left panel) and posterior (right panel) between TRENF and analytical expression for Gaussian random fields from the test set. . . . . 129
7.5 The comparison of the learned convolutional kernel $\tilde{T}(k)$ (top panel) and non-linearity $\Psi_y(x)$ (bottom panel) with the optimal solution (Equation 7.16 and 7.17) for GRFs. . . 129
7.6 The posteriors from TRENF (green contour) and power spectrum (red contour) on uncurated test data. . . . . 131
7.7 Comparison of the posterior constraints between generative learning (red contour) and discriminative learning (green contour) for the outlier cases in generative learning. . . . 132
7.8 A visualization of the simplified mask we considered in this work (left panel), and the data after the affine coupling (inpainting) layer (right panel). . . . . 133
7.9 The posterior distribution from TRENF on test data with mask. . . . . 134
7.10 Left panel: the loss curve of TRENF with $M = 0$ (spherical kernel) and $M = 2, N = 16$ on validation set. Right panel: the amplitude ratio $\tilde{A}_2(k)$ / $\tilde{A}_0(k)$ of the convolution kernels in different layers of TRENF with $M = 2, N = 16$. . . . . 136

8.1 Illustration of Multiscale Flow model. . . . . 145
8.2 Percentage of test data that fall outside $95\%$ confidence region for different $\lambda$ values. . . 147
8.3 Multiscale Flow posterior comparison of different scales on 20 test data with galaxy number density $n_g = 30\mathrm{arcmin}^{-2}$. . . . . 148
8.4 Comparison of posterior distributions between different scales of Multiscale Flow and power spectrum on a $3.5 \times 3.5\mathrm{deg}^2$ convergence map with $n_g = 20\mathrm{arcmin}^{-2}$. . . . . 150

8.5 Top panel: scale-dependent posterior analysis of a baryon-corrected convergence map using Multiscale Flow trained on dark-matter-only maps (left), and Multiscale Flow trained on BCM maps (right). Bottom panel: ROC curve of identifying distribution shift with $\log p$ (left) and $\Delta \log p$ (right). . . . . 152
8.6 Illustration of Multiscale Flow sample generation (the reverse of Figure 8.1). . . . . 153
8.7 Comparison of the power spectrum (left) and pixel probability distribution function (right) between simulations and Multiscale Flow samples at fiducial cosmology. . . . . 154

9.1 Comparison of the constraining power of different losses on 16 test data with the panels corresponding to galaxy number density $n_g = 10, 30, 100\,\text{arcmin}^{-2}$ respectively. . . . . 172
9.2 Comparison of the constraining power of different losses on 16 test data with the panels corresponding to galaxy number density $n_g = 20, 50, 100\,\text{arcmin}^{-2}$ respectively in the presence of baryons. . . . . 173

# List of Tables

4.1 List of MADLens simulation parameters that can be set by the user and their typical values. . . . . 60

5.1 The numerical parameters of LDL hybrid simulations, low resolution TNG300-3 and the target TNG300-1 hydrodynamical simulations . . . . . 83
5.2 The LDL architecture for predicting different baryon observables . . . . . 94

8.1 Comparison of the constraining power between different methods. . . . . 149
8.2 Similar to Table 8.1, with baryonic effects. . . . . 160
8.3 Empirical coverage probability of posteriors from different methods, after marginalizing over baryon parameters. . . . . 161

9.1 Comparison of the constraining power between different methods at fiducial cosmology. 172
9.2 Comparison of the constraining power using different map processing methodology at $n_g = 30$ arcmin$^{-2}$. . . . . 173
9.3 Comparison of the $1\sigma$ constraints of cosmological parameters between different methods at fiducial cosmology. . . . . 174
9.4 Comparison of the constraining power between different methods in the presence of baryons. . . . . 175
9.5 Comparison of the constraining power between different methods using Gaussian random maps transfer learning. . . . . 176
9.6 Comparison of the constraining power between different methods using lognormal maps transfer learning. . . . . 177

# Acknowledgments

Completing this dissertation has been a long journey, and I am deeply indebted to the many people who have supported me along the way.

First and foremost, I would like to express my sincere gratitude to my advisor, Uroš Seljak, for his unwavering guidance, support, and encouragement throughout my PhD journey. Uroš truly cares for my academic success and personal growth, providing invaluable support in my pursuit of an academic career. His insightful feedback has consistently sharpened my work, and his steadfast support has been a constant source of strength. Uroš encouraged me to build confidence in myself, and I will be forever inspired by his passion for science and his optimism in the face of challenges. I am profoundly grateful for the opportunity to learn from him and to work under his mentorship.

I would like to extend my heartfelt thanks to Rachel Mandelbaum, Josh Bloom, and Yu Feng. They have provided great help and advice on our collaborated projects, and I always appreciate the fruitful science discussions with them. I would also like to thank the other members of my dissertation and qualifying exam committee, Liang Dai and Saul Perlmutter, for their valuable input, constructive criticism, and generous investment of time.

I also want to thank all the incredible people that I had the privilege of meeting and interacting with during my six years at the Berkeley Center for Cosmological Physics: Martin White, Lingyuan Ji, Minas Karamanis, Rongpu Zhou, Jia Liu, Simone Ferraro, Francois Lanusse, Satya Gontcho A Gontcho, Chirag Modi, Henry Liu, Stephen Chen, Jakob Robniki, and Noah Sailer. I am also deeply grateful to my wonderful collaborators during graduate school: Francisco Villaescusa-Navarro, Sukhdeep Singh, George Stein, Richard Grumitt, Vanessa Boehm, and Divij Sharma. Special thanks to James Sullivan, Adrian Bayer, Byeonghee Yu, Yin Li, and Hongming Zhu - the chats, discussions, and shared experiences have made the challenging moments of this journey bearable and the successes even sweeter.

I owe thanks to the teachers and professors who have guided and inspired me throughout my journey. I am especially thankful to Jichao Hu, who first ignited my passion for physics in high school, and to Zuhui Fan, who introduced me to the fascinating world of cosmology. I am also profoundly grateful to Brant Robertson and Yi Mao, under whose mentorship I completed several enjoyable and formative research projects during my undergraduate years. These experiences solidified my decision to pursue a PhD and have been pivotal in shaping my academic path.

My deepest thanks go to my friend and housemate Dun Tang, the most generous and kind-hearted person I have ever met. I feel incredibly fortunate to have shared so much of my life's journey with Dun — we were classmates in middle school and high school, and we continued to the same university and graduate school. His company and friendship have been a constant source of comfort and joy. I would also like to thank my friends in Berkeley, including Tianye Wang, Keming Zhang, Jiahao Yao, and Ziwen Zhao, for their companionship throughout these years. I also want to thank my old friends who kept in touch all these years, especially Zhili Zhang and Hao Zhang. Our shared experiences have made this journey all the more meaningful. Finally, I would like to thank my friends at Carnegie Mellon University, where I spent significant time during my PhD: Xiangchong Li, Yueying Ni, Tianqing Zhang, Hung-Jin Huang, Alan Zhou, and Yihao Zhou.

To my parents, your love and support have been the foundation upon which I have built my academic career. Thank you for always believing in me and providing constant support. To my partner, Nianyi Chen, your understanding, support, and encouragement have been my greatest source of strength. I am deeply thankful for your love and for the countless ways you have helped me navigate the challenges of this process. Meeting you is the luckiest thing that has happened to me.

This dissertation is as much a testament to the support and contributions of these individuals as it is a reflection of my efforts. Thank you all for being part of this journey.

# Chapter 1

# Introduction

Over the past few decades, cosmology has transitioned into a precision science thanks to the abundant cosmological observations, revolutionizing our understanding of the Universe. At the heart of modern cosmology lies the standard cosmological model, often referred to as the $\Lambda\mathrm{CDM}$ model. This model provides a comprehensive framework for interpreting a vast array of observational data, ranging from the cosmic microwave background (CMB) to the large-scale distribution of galaxies. The $\Lambda\mathrm{CDM}$ model is grounded in the framework of General Relativity and the cosmological principle, which holds that the universe is homogeneous and isotropic on large scales. It combines two crucial components: dark energy, represented by the cosmological constant ($\Lambda$), and cold dark matter (CDM). Dark energy accounts for approximately 70% of the total energy density of the universe [257] and is thought to be responsible for the observed accelerated expansion. Cold dark matter, making up about 25% of the universe's mass-energy content, interacts gravitationally but not electromagnetically, influencing the formation and clustering of galaxies and large-scale structures. The remaining 5% consists of ordinary baryonic matter, including stars, planets, and interstellar gas.

According to the $\Lambda\mathrm{CDM}$ model, the universe began approximately 13.8 billion years ago with the Big Bang, followed by a brief period of exponential expansion known as inflation. This rapid expansion smoothed out any initial irregularities and set the stage for the formation of large-scale structures. Around 380,000 years after the Big Bang (around redshift 1100), the universe cooled sufficiently for protons and electrons to combine and form neutral hydrogen atoms. This recombination event made the universe transparent to radiation, resulting in the release of the CMB, which we observe today as a relic of this early epoch. Following recombination, the universe entered a period known as the dark ages, characterized by the absence of luminous sources. The formation of the first stars and galaxies several hundred million years later initiated the reionization era, where photons from these early objects reionized the intergalactic medium. Over billions of years, small density fluctuations seeded by inflation grew under the influence of gravity, leading to the formation of galaxies, clusters, and the cosmic web. Cold dark matter played a crucial role in this process by providing the gravitational potential in which baryonic matter could coalesce. About 4 or 5 billion years ago (redshift 0.4), the influence of dark energy began to dominate, causing the expansion of the universe to accelerate. This acceleration continues to shape the large-scale

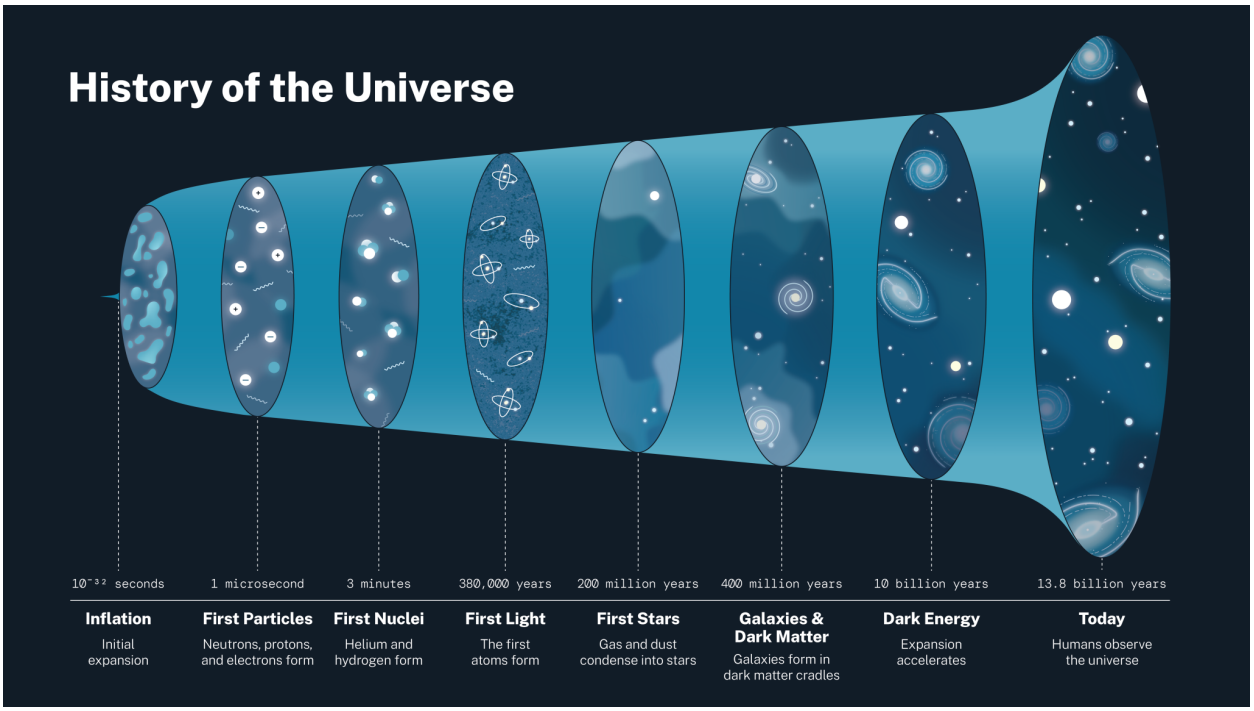



Figure 1.1: The cosmic history according to the $\Lambda$CDM model from Big Bang to the present day. Credit: NASA.

dynamics and fate of the cosmos. A visualization of the cosmic history is shown in Fig. 1.1.

The $\Lambda$CDM model has been remarkably successful in explaining a wide range of cosmological observations. Its predictions are consistent with the CMB data from Planck [257], the large-scale structure observed by, e.g., SDSS [25], DES [2] and HSC [3], the cosmic acceleration inferred from supernovae, and the observed abundances of light elements. However, several tensions and mysteries challenge the model and prompt ongoing research and debate, including the inconsistent measurement between the early universe and late-time universe on the expansion rate of the universe $H_0$ (the "Hubble tension") [271] and the amplitude of matter fluctuations $S_8$ (the "$S_8$ tension") [130, 2, 183]. Moreover, the nature of the two building blocks of $\Lambda$CDM, dark matter, and dark energy, remains unknown. The extensive searches for dark matter particles have so far ruled out much of the parameter space but still shown no detection, and for dark energy constraints, recent LSS analysis has shown deviations from the cosmological constant model [65].

The next generation of cosmological surveys, known as stage IV surveys, promises to address these tensions and mysteries and provide unprecedented insights into the universe's structure and evolution. These ambitious projects include DESI [66], Euclid[9], SPHEREX [74], Rubin Observatory LSST [195], Roman Space Telescope [75], and CMB S4 [1]. These surveys will provide unprecedented amounts of data across different cosmological probes. It is thus an important task to build the theory and analysis tools to optimally extract cosmological information and fully

realize the potential of these vast datasets, which is the main focus of this dissertation.

## 1.1 Large-Scale Structures

The matter is not randomly or uniformly distributed in the universe. On large scales it is clustered and forms structures like halos, filaments, and voids, a complex and vast network known as the cosmic web, or large-scale structure (LSS). The LSS is seeded by small quantum fluctuations in the early Universe and grows under gravitational evolution and cosmic expansion, therefore it contains significant information about primordial physics, the gravity law, the matter and energy content of our universe, and its accelerated expansion. Studying LSS allows us to put strong constraints on our current cosmological model, to understand its mysterious components such as the nature of dark energy, and to explore new physics such as modified gravity. To model the LSS and its evolution, let us first define the density contrast $\delta$ as

$$\delta(x, a) = \frac{\rho(x, a)}{\bar{\rho}(a)} - 1, \tag{1.1}$$

where $\rho(x, a)$ is the matter density at comoving coordinate $x$ and scale factor $a$. The gravitational potential $\phi$ sourced by the matter density fluctuation $\delta_m$ is given by the Poisson equation

$$\nabla^2\phi(x, a) = \frac{3}{2}\mathcal{H}^2(a)\Omega_m(a)\delta_m(x, a), \tag{1.2}$$

where $\mathcal{H}$ is the conformal Hubble parameter and is related to Hubble parameter $H(t) = \dot{a}(t)/a(t)$ as $\mathcal{H} = aH$, $\Omega_m(a) = \rho_m(a)/\rho_c$ is the mean matter density in the universe and $\rho_c(a) = \frac{3H^2(a)}{8\pi G}$ is the critical density.

Now let us adopt the Lagrangian description of the LSS, and consider a particle at comoving coordinate $x$. Its equation of motion in the Newtonian limit is given by

$$\ddot{x} + \mathcal{H}(\tau)\dot{x} = -\nabla\phi, \tag{1.3}$$

where the derivative is taken with respect to the conformal time $\tau = a(t)t$. Combining the Poisson equation 1.2 and the equation of motion 1.3, and with the cosmic background evolution $a(t)$ given by Friedmann equations, one can solve for the evolution of LSS.

This set of differential equations can be solved analytically with perturbation theory (i.e., Lagrangian perturbation theory, LPT). In LPT, one solves the differential equations perturbatively by expanding the particle displacement field $\Psi(q, \tau)$ successively to different orders

$$\Psi(q, \tau) = x(\tau) - q = \Psi^{(1)}(q, \tau) + \Psi^{(2)}(q, \tau) + \cdots, \tag{1.4}$$

where $q$ is the Lagrangian (comoving initial) coordinate. While perturbation theories work well on large scales, they are not able to model shell crossing and the performance worsens as we go to non-linear scales. To accurately model the LSS down to small scales, one normally solves the dynamics numerically with cosmological simulations, which are discussed in the next section.

## 1.2 Cosmological Simulations

Cosmological simulations are computational models used to simulate the evolution of the Universe in a periodic box. They provide a bridge between theoretical models and observational data, allowing for a detailed and quantitative understanding and prediction of the LSS we observed in the universe. Here we introduce the basics of cosmological simulations, focusing on gravity-only N-body simulations, which enables many works done in this dissertation.

N-body simulations use discrete particles as tracers for the matter distribution. The initial condition (IC) of these particles is often set with first [362] or second [151, 99], or third-order LPT [221] at an early enough time when the gravity evolution is only mildly nonlinear and can be well described by low-order perturbation theories. The particles are then evolved under gravitational potential $\phi$ by numerically integrating the equation of motion (equation 1.3).

A popular method for estimating the gravitational potential $\phi$ is the particle-mesh (PM) approach, which solves the Poisson equation (equation 1.2) using a mesh-based method. The particles are firstly painted on a mesh to estimate the overdensity field $\delta$, and then the discretized Poisson equation can be solved efficiently in Fourier space using fast Fourier transform. The PM method is computationally efficient ($O[\mathcal{N}\log(\mathcal{N})]$), where $\mathcal{N}$ is the mesh size) and can be easily parallelized, but the force accuracy is fixed at the grid size.

To estimate the gravitational force at scales below the grid size, one can use the particle-particle (PP) method to solve the Poisson equation in its integral form:

$$\phi(\mathbf{x}_i) \propto -\sum_{j=1}^{N} \left\{ \frac{m_j}{|\mathbf{x}_j - \mathbf{x}_i| + \epsilon\left(|\mathbf{x}_j - \mathbf{x}_i|\right)} \right\}. \tag{1.5}$$

Here $\epsilon\left(|\mathbf{x}_j - \mathbf{x}_i|\right)$ is a softening kernel that prevents the singularity at close separations. P$^3$M combines the PM approach with the direct PP method, by computing the short-range gravity from direct PP summation, and computing long-range gravity with PM. The PP summation can also be approximated by the oct-tree algorithm [15] to reduce its computational cost, leading to the TreePM method [13].

Numerically integrating the equation of motion (equation 1.3) usually takes hundreds to thousands of time steps and can be computationally expensive. Several works have tried to reduce the number of time steps while maintaining accuracy at large scales, by incorporating the perturbation theories into the N-body solvers [89, 189, 312]. In a leapfrog numerical integrator, the particle positions $x$ and velocities $p$ are updated with [261]

$$x(a_1) = x(a_0) + \mathcal{D}p(a_r) \tag{1.6}$$
$$p(a_1) = p(a_0) + \mathcal{K}f(a_r), \tag{1.7}$$

where $f$ is the force, $a_0$ is the initial scale factor, $a_1$ the final scale factor, $a_r$ the reference scale factor, and $\mathcal{D}$ and $\mathcal{K}$ are drift and kick factors. The basic idea of FastPM [89] and perturbation theory-informed integrators [189] is to modify the drift and kick factors to be compatible with LPT, such that the large-scale matter distribution can be predicted accurately with only a few time steps.

However, as discussed in the previous section, perturbation theory breaks down at the scale of shell crossing, and these perturbation theory-informed integrators begin to lose power at non-linear scales when the number of time steps is reduced. In chapter 2, 3 and 4 we develop numerical models to compensate for this effect and improve the small-scale modeling.

So far we have focused on modeling the evolution of dark matter and ignored the baryonic physics. Hydrodynamical simulations also model the baryon component, with additional physics processes including gas pressure, cooling, heating, star formation, stellar feedback, massive black hole seeding, accretion, feedback, and dynamics. Due to the finite resolution of cosmological simulations, most (if not all) of these baryonic processes have to be simulated effectively with subgrid models. While being most realistic, hydrodynamical simulations are computationally too expensive to run in cosmological volumes comparable to current LSS surveys. Empirical models such as halo occupation distribution (HOD) [21], abundance matching [322], and baryonic correction models [281] are widely used to model baryon observables from N-body simulations. In chapter 5 and 6, we also introduce effective machine learning models that connect dark matter from N-body simulations to baryons with low computational cost.

## 1.3 Cosmological Inference

Cosmology relies heavily on Bayesian inference to interpret complex observational data and draw meaningful conclusions about the universe and the underlying physical law. Bayesian inference is a powerful statistical framework that provides a coherent and principled approach to updating our knowledge about unknown model parameters based on observed data. It is based on Bayes' theorem, which produces the posterior probability distribution of model parameter $\theta$, given the observed data $D$ (e.g., some tracer field of LSS) and the underlying model assumption $M$

$$p(\theta|D, M) = \frac{p(D|\theta, M)p(\theta|M)}{p(D|M)}, \tag{1.8}$$

where $p(D|M) = \int p(D|\theta, M)p(\theta|M)d\theta$ is a normalization and also known as Bayesian evidence, $p(\theta|M)$ is the prior, and $p(D|\theta, M)$ is the likelihood function. While Bayesian evidence can be important in model comparison, in parameter inference we are mostly interested in estimating posterior with Monte Carlo sampling and the evidence can be ignored. We can also drop the dependence on model $M$ since in parameter inference we normally consider a single model (e.g., $\Lambda$ CDM).

The prior function encodes the degree of belief about the values of $\theta$ before the observation. Its choice can be important when the likelihood is not very informative and when there is a large number of nuisance parameters that correlate with the parameter of interest. In these cases, even the naive uniform prior can lead to significant bias in parameter constraints, which is also known as the prior volume effect, and has been observed in LSS analysis with effective field theories (EFT) [293, 40]. For field-level simulation-based inference (SBI, discussed in the next section), this is not likely to be a problem, since the field-level likelihood is much more informative, and SBI usually comes with much fewer nuisance parameters than EFT.

The likelihood function $p(D|\theta)$ is of central importance in statistical inference. In the LSS analysis, however, the exact likelihood function is intractable due to nonlinear gravity evolution and baryonic process. In traditional LSS analysis, cosmologists usually make two simplifications to the likelihood function. Firstly, the observed high-dimensional tracer field (or catalog) $D$ is compressed into low-dimensional summary statistics $S(D)$, and likelihood is defined on the summary statistics $p(S(D)|\theta)$. A natural choice of $S(D)$ is the two-point correlation function $\xi(r)$

$$\xi(r) = \langle \delta(x)\delta(x+r) \rangle. \tag{1.9}$$

It describes the excess probability of finding two galaxies or tracers separated by a distance $r$. The Fourier analog of the two-point correlation function, the power spectrum $P(k)$, is defined by averaging over Fourier modes $\delta(k)$

$$(2\pi)^3 P(k)\delta_D(k-k') = \langle \delta(k)\delta^*(k') \rangle. \tag{1.10}$$

While the two-point statistics are natural choices even in the nonlinear regime, they do not capture all the information in the LSS data, and adding higher-order information is less straightforward. Even adding the three-point function means adding a function of three parameters, which is considerably more complex to describe than the two-point function. Numerous other ad-hoc statistics have been proposed, from peak counts to void counts, void profiles, etc.. Recently, machine learning algorithms are also applied to construct informative summary statistics[41, 270, 202]. While they have all been shown to provide complementary information to the power spectrum, it is not clear if any or the combination of all of these statistics can fully capture the full information content in the data, and how to design the most informative summary statistics is still a popular research topic.

For a given summary statistic $S(D)$, the likelihood function $p(S(D)|\theta)$ is usually approximated by a multivariate Gaussian distribution

$$p(S(D)|\theta) = \frac{1}{\sqrt{(2\pi)^n |C|}} \exp\left( -\frac{1}{2}(S(D) - S(\theta))^T C^{-1} (S(D) - S(\theta)) \right), \tag{1.11}$$

where the mean expectation $S(\theta)$ needs to be evaluated as a function of cosmological and nuisance parameters $\theta$. For two-point statistics, this can be a good approximation due to the central limit theorem, but it may not hold for some non-gaussian statistics [86].

An alternative approach is using the reconstruction of initial conditions and performing the marginal integral over these latent variables so that we are left with marginal data likelihood as a function of cosmological parameters $p(x|y)$:

$$p(x|y) = \int p(x|y,z)p(z|y)dz, \tag{1.12}$$

where $z$ stands for the initial conditions, $p(z|y)$ is a Gaussian distribution fully specified by the linear power spectrum, and $p(x|y,z)$ involves running a simulation to predict the noiseless data, and a noise model which is usually approximated by a Gaussian. In this approach, the inference is performed at the field level, and in principle it is able to extract all the information in the data.

However, performing this marginal integral over initial conditions $z$ has proven to be difficult. One can sample over $z$ and $y$ using Hamiltonian Monte Carlo (HMC) [147, 163, 346], but the samples are usually very correlated in high dimensions even if HMC is used, and thousands of full N-body simulation steps may be needed between two independent samples. An alternative to this marginalization approach is maximum a posteriori (MAP) estimation of $z$, followed by an MAP-based unbiased estimator of the cosmological parameters [287, 224]. While obtaining the MAP of $z$ is faster, making the estimate unbiased and obtaining the cosmological parameter posterior is still expensive and can be suboptimal.

## 1.4 Simulation-Based Inference

Simulation-based inference (SBI, sometimes also known as likelihood-free inference) [51] uses numerical simulations to perform statistical parameter inference without relying on analytical likelihood function assumptions like Equation 1.11. Many different approaches belong to the SBI family, ranging from Approximate Bayesian Computation (ABC), amortized likelihood approach, amortized posterior approach, amortized likelihood ratio approach, etc. Strictly speaking, the forward modeling reconstruction method discussed in the previous section is also a form of SBI, where the simulators are used directly during inference [51].

In this dissertation, we will focus on the amortized likelihood approach (i.e., neural likelihood estimation, NLE) and amortized posterior approach (i.e., neural posterior estimation, NPE). In both cases, one firstly runs a large number of cosmological simulations to generate mock data $x$ with random realizations and cosmological parameters $\theta$ drawn from the prior (For NPE, this prior has to be the same as the prior in parameter inference. For NLE, it does not have to be). Then from this set of $\{x_i, \theta_i\}$ pair, one tries to estimate the conditional probability distribution $p(x|\theta)$ (NLE) or $p(\theta|x)$ (NPE) using density estimation machine learning models such as normalizing flows (introduced below). The learned likelihood function $p(x|\theta)$ or posterior $p(\theta|x)$ can then be used for parameter inference given observation data.

Normalizing Flows play a central role in SBI. They are powerful frameworks for density estimation [71, 239] and sampling [161]. These models map the data $x$ to latent variables $z$ through a sequence of invertible transformations $f = f_1 \circ f_2 \circ ... \circ f_n$, such that $z = f(x)$ and $z$ is mapped to a base distribution $\pi(z)$. The base distribution $\pi(z)$ is normally chosen to be a standard normal distribution, i.e. a Gaussian white noise with zero mean and unit variance, $\pi(z) = \mathcal{N}(0, I)$. The probability density of data $x$ can be evaluated using the change of variables formula:

$$\begin{aligned} p(x) = & \quad \pi(f(x)) \left|\det\left(\frac{\partial f(x)}{\partial x}\right)\right| \\ = & \quad \pi(f(x)) \prod_{l=1}^{n} \left|\det\left(\frac{\partial f_l(x)}{\partial x}\right)\right|. \end{aligned} \tag{1.13}$$

To sample from $p(x)$, one first samples latent variable $z$ from $\pi(z)$, and then transform variable $z$ to $x$ through $x = f^{-1}(z)$. The transformation $f$ is usually parametrized with neural networks $f_\phi$,

and the parameters $\phi$ are estimated using Maximum Likelihood Estimation (MLE):

$$\phi^* = \arg\max_{\phi} \frac{1}{N} \sum_{i=1}^{N} \log p_\phi(x_i), \tag{1.14}$$

where the data likelihood $p(x)$ is given by Equation 1.13. The MLE solution minimizes the Kullback-Leibler (KL) divergence between the model distribution $p_\phi(x)$ and the true data distribution. The parameterization of $f$ must satisfy the requirements that the Jacobian determinant $\det(\frac{\partial f_l(x)}{\partial x})$ is easy to compute for evaluating the density, and the transformation $f_l$ is easy to invert for efficient sampling.

SBI is promising as it leverages high-fidelity cosmological simulations and allows extracting non-Gaussian information from non-linear scales that cannot be modeled analytically. While it has been successfully applied to a few cosmological inference problems so far[114, 148, 100], it still faces several challenges that limit its applications:

- Distribution shift (model misspecification): If the simulations do not realistically represent our universe (due to, e.g., missing physics, inaccurate physical models, or not fully accounting for systematic effects), there would be a shift between the training distribution (simulations) and the test data (observations), known as distribution shift or model misspecification. Distribution shift could lead to significant bias in parameter inference [37].

- High computation cost of cosmological simulations: It is computationally challenging to run cosmological simulations that satisfy the following 3 requirements simultaneously: 1) with high enough resolution to avoid model misspecification (first point); 2) in a volume comparable to the upcoming cosmological surveys; and 3) with enough number of simulations to fully sample the parameter space for training the SBI model [225].

- Unfaithful posterior uncertainty: It has been suggested that all SBI methods could produce overconfident posterior constraints in some applications [128]. It is therefore crucial to validate that the trained SBI model is well-calibrated before applying it to real data.

- Lack of interpretability: Normalizing flows are based on neural networks, which are considered black boxes by many researchers in the community. For high-dimensional summary statistics and field-level SBI analysis, it is not clear where the model is extracting information from.

## 1.5 Dissertation Outline

As discussed in previous sections, SBI provides a promising framework for extracting rich non-Gaussian information from current and upcoming surveys. However, it faces several challenges that limit its applications. To facilitate the deployment of SBI approaches into upcoming survey analysis pipelines, in this dissertation, we try to answer the following two questions

- How can we simulate the LSS accurately with low computational cost, so that we can efficiently run many simulations across the parameter space for SBI training?
- How to design and train the SBI models for optimal and robust performance in cosmological applications?

We divide the dissertation into two parts to answer these two questions separately. In chapters 2 - 6, we focus on developing effective models to accurately simulate LSS with low computation cost. In chapter 2, we develop gradient-based models to improve the modeling of nonlinear gravity and baryonic feedback on small scales in low-resolution fast simulations. These models are incorporated into lightcone simulations in chapter 3 for efficient and accurate modeling of weak gravitational lensing. We also modify the halo finder to improve the predicted halo catalogs in fast simulations. In chapter 4, we rewrite our models and fast weak lensing simulations with differentiable programming and make them publicly available to the community.

We turn to the modeling of baryons in chapter 5 and 6. In chapter 5 we develop a Lagrangian Deep Learning (LDL) model for modeling various baryonic observables from N-body simulations. By combining N-body solver FastPM with LDL, we are able to predict a wide range of baryonic observables including stars, kinematic and thermal Sunyaev–Zeldovich effect, with nearly four orders of magnitude lower computation cost compared to the hydrodynamical simulations. In chapter 6, we find an efficient way to model the baryonic feedback on the total matter distribution, and build an emulator of baryonic effect at the field level using Gaussian processes. This method and our emulator enable field-level weak lensing SBI analyses with baryonic effects.

In the second half of this dissertation (chapter 7, 8 and 9), we focus on the development of field-level SBI models. For superior performance in cosmological applications, we integrate physics constraints and domain knowledge into ML models. In chapter 7, we develop a Translation and Rotation Equivariant Normalizing Flow (TRENF) model that satisfies symmetry constraints. We apply TRENF to learn the high-dimensional field-level data likelihood, and show that it leads to significant improvements in constraining power over the standard power spectrum summary statistic. In chapter 8, we introduce Multiscale Flow (MSF) for optimal and robust field-level analysis. MSF uses hierarchical decomposition of cosmological fields via a wavelet basis, and then models different wavelet components separately as Normalizing Flows. This decomposition allows us to separate the information from different scales and identify distribution shifts in the data such as unknown scale-dependent systematics, thus making our analysis robust to some distribution shifts. In chapter 9, we perform a comparative study and explore the impact of different analysis choices on parameter constraints for Convolutional Neural Network (CNN) based SBI models. By performing a large number of control experiments, we identify the best hyperparameter choice and training loss functions that improve on previous CNN-based SBI models by a factor of 2. Finally, we conclude in chapter 10.

# Chapter 2

# A gradient-based method for modeling baryons and matter in halos of fast simulations

The contents of this chapter were originally published in [53],

> A gradient based method for modeling baryons and matter in halos of fast simulations
> *Dai B., Feng Y., Seljak U.*(arXiv:1804.00671) JCAP 11 (2018) 009

Fast N-body PM simulations with a small number of time steps such as FastPM or COLA have been remarkably successful in modeling the galaxy statistics, but their lack of small-scale force resolution and long-time steps cannot give accurate halo matter profiles or matter power spectrum. High-resolution N-body simulations can improve on this, but lack baryonic effects, which can only be properly included in hydro simulations. In this chapter, we present a scheme to calibrate the fast simulations to mimic the precision of the hydrodynamic simulations or high-resolution N-body simulations. The scheme is based on a gradient descent of either effective gravitational potential, which mimics the short-range force, or of effective enthalpy, which mimics gas hydrodynamics and feedback. The scheme is fast and differentiable, and can be incorporated as a post-processing step into any simulation. It gives very good results for the matter power spectrum for several of the baryonic feedback and dark matter simulations, and also gives improved dark matter halo profiles. The scheme is even able to find the large subhalos, and increase the correlation coefficient between the fast simulations and the high-resolution N-body or hydro simulations. It can also be used to add baryonic effects to the high-resolution N-body simulations. While the method has free parameters that can be calibrated on various simulations, they can also be viewed as astrophysical nuisance parameters describing baryonic effects that can be marginalized during the data analysis. In this view, these parameters can be viewed as an efficient parametrization of baryonic effects.

## 2.1 Introduction

Extracting accurate cosmological information from the current and future sky surveys requires high-precision simulations. Computer simulations with quasi N-body numerical schemes provide an alternative to the full N-body or hydro simulations for creating fast realizations of the large-scale structure (LSS), but lack resolution on small scales. Methods in this family includes FastPM [89], COLA [312], subcycle TreePM [307], and PPM-GLAM [168].

Although the halo catalog from the quasi N-body simulations is well correlated to a true N-body simulation of the same initial condition [89], the dark matter density distribution is less accurate, since it requires not just the center of mass of a dark matter halo but also its density profile. For example, a 10 step FastPM N-body simulation misses about $15\% - 20\%$ of power at the Nyquist frequency of the PM grid [89, 312]. Accurate modeling of the dark matter requires more time steps. For example, Izard et al. [142] proposed ICE-COLA, in which the code parameters are optimized to achieve a matter power spectrum within 1 percent for $k \lesssim 1\mathrm{hMpc}^{-1}$, at the cost of using 40-time steps and a force mesh that is 3 times smaller than the particle mean separation. This leads to about an order of magnitude higher cost than a 10-step simulation without a higher force mesh resolution.

The suppression in the power spectrum is due to the inability of these methods to resolve the internal structures of halos. With such large time steps, the halo is not fully virialized, resulting in a much shallower density profile than the NFW profile [230, 229]. This severely undermines the application of quasi-N-body simulations to the applications where modeling underlying dark matter fields is crucial, including SZ effects and weak lensing.

On the other hand, even the most accurate N-body simulation cannot model the effect of baryons without introducing costly and complex numerical schemes to model the hydrodynamics, cooling, star formation, and AGN feedback. The effects of baryon on the density profile are also important. Gravity leads to collapse, but baryons, due to the gas pressure, resist the collapse. In addition, the feedback from AGN and supernovae can transport a large amount of gas to the outskirts of the halos. This can even lead to an expansion of the dark matter halo and the reduction of the halo mass [79, 317, 214, 328]. It has been shown that these effects can reduce the matter power spectrum by more than ten percent for $1 \lesssim k \lesssim 30\mathrm{hMpc}^{-1}$ [326, 342].

In this chapter, we introduce a simple numerical scheme that allows an arbitrary calibration of the dark matter density field against high-resolution N-body simulations and hydrodynamical simulations. Our method is based on the motion of particles along the gradient direction of a scalar field that is generated from the existing density field. While keeping the large-scale structure unchanged, our model adds additional displacements to the particles to help resolve the small-scale structures in quasi-N-body simulations, or to effectively model the baryonic feedback. This spirit is similar to the Augmented Lagrangian Perturbation Theory (ALPT) [164] and the baryonic correction model [281]. Other similar methods include one-point remapping [180] and MUSCLE [233]. However, ALPT assumed the analytical spherical collapse model and determined the small-scale displacement with local density, while our method is based on a PM solver and solves a modified potential. Our method is more flexible in terms of matching to high-resolution simulations (see Section 2.3).

Our method is simple in the sense that the Jacobian of the method has a very simple form,

and can be easily embedded into a parameter inference framework. It is essentially data-driven learning and can be viewed as a form of machine learning (ML): rather than having a full dynamical model for the matter density distribution, we train the low-resolution simulation to reproduce the results of the more expensive high-resolution simulations. Training is performed on the comparison between the low and high-resolution simulations. However, our specific approach is gradient-based rather than using standard ML techniques, which we argue offers several advantages, primarily low dimensionality of parameter space to be optimized against.

The plan of the chapter is as follows: we first describe the mathematics and motivation of the numerical schemes in Section 2.2. In Section 2.3 we show the performance of the gradient-based schemes in an example, with the emphasis on matter power spectra, halo profiles, and sub-halo statistics. We present a recommendation of the parameter choices in Section 2.4. Finally, we conclude in Section 2.5.

## 2.2 Gradient-based learning: theory and motivation

In this section, we introduce the gradient-based method and derive several forms depending on the problem we wish to solve. The basic goal is that we would like to have a scheme that mimics the physics that is missing in low-resolution simulations. This can be either short-range force in the case of gravity, or hydrodynamic force effects in the case of hydrodynamic simulations. Our main idea is that instead of performing a full simulation we can partially account for these effects as a post-processing step, which will perform the effective missing short-range force operation only once. Since force is a gradient of some effective potential this leads to the idea of descending along its gradient. The parameters controlling the model can be learned from the high-resolution simulations themselves, or more precisely by comparing high and low-resolution simulations. This is a form of data-driven learning, but since it is based on some notion of missing physics it can be parametrized with relatively few free parameters. We describe several versions of this idea below.

### Descent along the gradient of gravitational potential: Particle-Particle interaction

Without the short-range particle-particle force, quasi-nbody schemes such as FastPM[89] do not resolve the structures with scales smaller than the mesh resolution, resulting in halo profiles that are shallower than their full N-body counterparts. A straightforward way to improve this would be adding a short-range Particle-Particle (PP) force in every step during the simulation, which would however also induce a large additional computational cost. However, the effect of this short-range PP force is to aid the collapse of halos, which is a radial motion of the particles. We thus propose a simple model in which we move the particles along a short-range PP force after the simulation has finished. The displacement bypasses the momentum of the particles, corresponding to the case of a fluid with infinite viscosity. The direction of the short-range PP force points towards the potential minimum. Therefore, displacing particles along this direction can sharpen the profile of halos.

A tree can be built to accelerate the calculation summation of short-range PP force,

$$\mathbf{F}_{\mathsf{PP}} = \sum_{r<r_{max}} \frac{Gm_1m_2}{(r+\epsilon)^3} \cdot \mathbf{r} \tag{2.1}$$

where $r_{max}$ determines the largest separation of particle pairs to calculate the force, and $\epsilon$ is the force softening. The KDTree implementation in kdcount is improved to compute the short-range force [90].

The displacement is proportional to the force and given by

$$\mathbf{S} = (\alpha_{\mathsf{PP}}/H_0^2)\, \mathbf{F}_{\mathsf{PP}} \tag{2.2}$$

where $\alpha_{PP}$ is a free parameter in our model and $H_0$ is the Hubble parameter induced here to make $\alpha_{\mathsf{PP}}$ dimensionless.

We can vary the 3 parameters ($\alpha_{PP}$, $\epsilon$, and $r_{max}$) to match the power spectrum against N-body and hydrodynamical simulations or against averaged halo profiles. This is discussed further in section 2.7.

**Descent along the gradient of gravitational potential: Particle-Mesh interaction**

Particle-Mesh (PM) force is significantly cheaper to compute than a PP force. We therefore consider a faster scheme where the short-range force is instead computed with a Fourier space particle mesh solver. The full gravitational potential of a particle mesh solver is given by

$$\phi = 4\pi G\bar{\rho}\nabla^{-2}\delta, \tag{2.3}$$

where the force is given by the gradient of the potential.

Drifting the particles along the gradient of the gravitational potential will act as an additional time step, increasing the large-scale growth. We can eliminate the large-scale component of the potential with a high pass filter $\widehat{O_1}$,

$$\hat{\mathbf{O}}_{\mathbf{l}}(k) = \exp\left(-\frac{k_l^2}{k^2}\right) \tag{2.4}$$

where $k_l$ is the long range scale parameter. The spirit of the high pass filter is similar to the $r_{max}$ parameter in our PP model: this filter removes the long-range force by damping the potential modes with scales larger than $k_l$. On small scales, to reduce the numerical effect induced by the mesh resolutions, we introduce another low-pass filter

$$\hat{\mathbf{O}}_{\mathbf{s}}(k) = \exp\left(-\frac{k^4}{k_s^4}\right) \tag{2.5}$$

where $k_s$ is the short range scale parameter. The low pass filter has a similar effect as force softening $\epsilon$, although the cut-off is slightly sharper. We show an example of the filter in Figure 2.1.

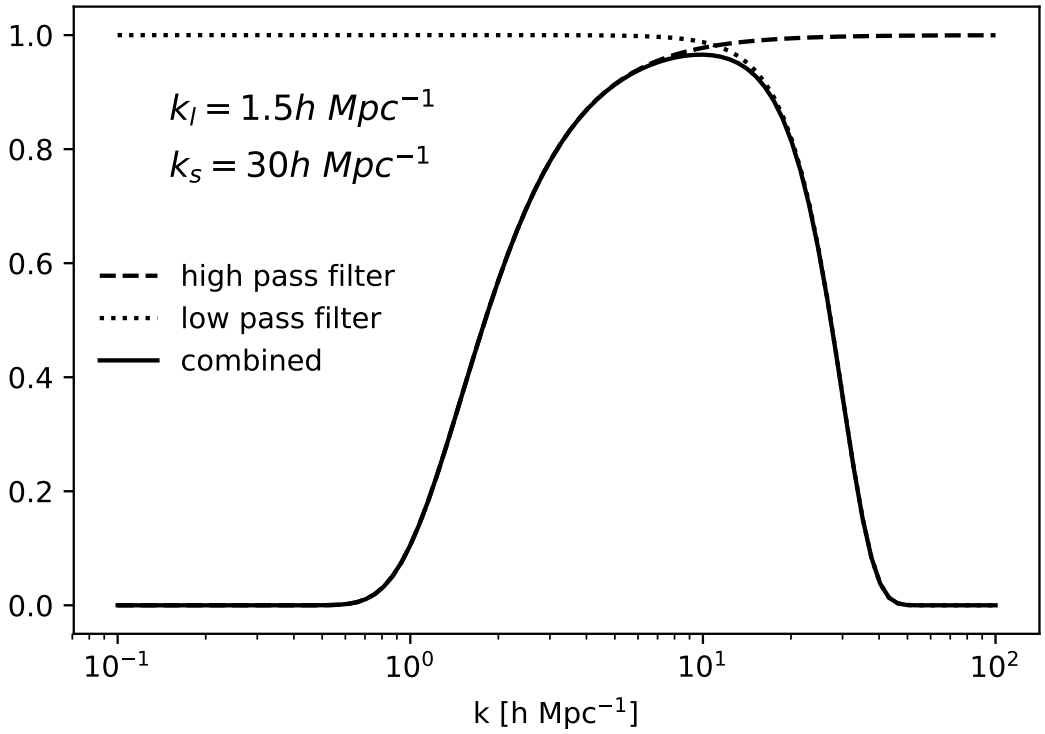



Figure 2.1: The filter as a function of k.

We compute the displacement with the filtered potential:

$$\mathbf{S} = (\alpha_{\mathsf{PM}}/H_0^2)\ \mathbf{F}_{\mathsf{PM}} \tag{2.6}$$
$$= (\alpha_{\mathsf{PM}}/H_0^2)\ \nabla\hat{\mathbf{O}}_{\mathbf{l}}\hat{\mathbf{O}}_{\mathbf{s}}\phi \tag{2.7}$$
$$= (4\pi G\bar{\rho}\alpha_{\mathsf{PM}}/H_0^2)\ \nabla\hat{\mathbf{O}}_{\mathbf{l}}\hat{\mathbf{O}}_{\mathbf{s}}\nabla^{-2}\delta \tag{2.8}$$

In Figure 2.2, we show that Equation 2.8 achieves a similar effect comparing Equation 2.2, but the former one is much faster, suggesting that a full resolution PP force solver is unnecessary. We will only show the results from the PM method in the rest of this chapter. We name this scheme the Potential Gradient Descent model (hereafter PGD).

In figure 2.3 we show how the parameters influence the matter power spectrum. Here we focus on the parameter $\alpha$ in the PGD model fixing the parameters $k_l$ and $k_s$. As expected, as $\alpha$ increases, the halo profiles are sharpened and the small-scale power spectrum increases. When $\alpha \gtrsim 0.015$, the particles begin passing the halo centers, and increasing $\alpha$ will smooth the density field, making the small-scale power spectrum decrease. The maximum power enhancement is achieved at $\alpha = 0.015$ (depending on $k_l$ and $k_s$). We see from the figure that varying $\alpha$ can match both the effects of

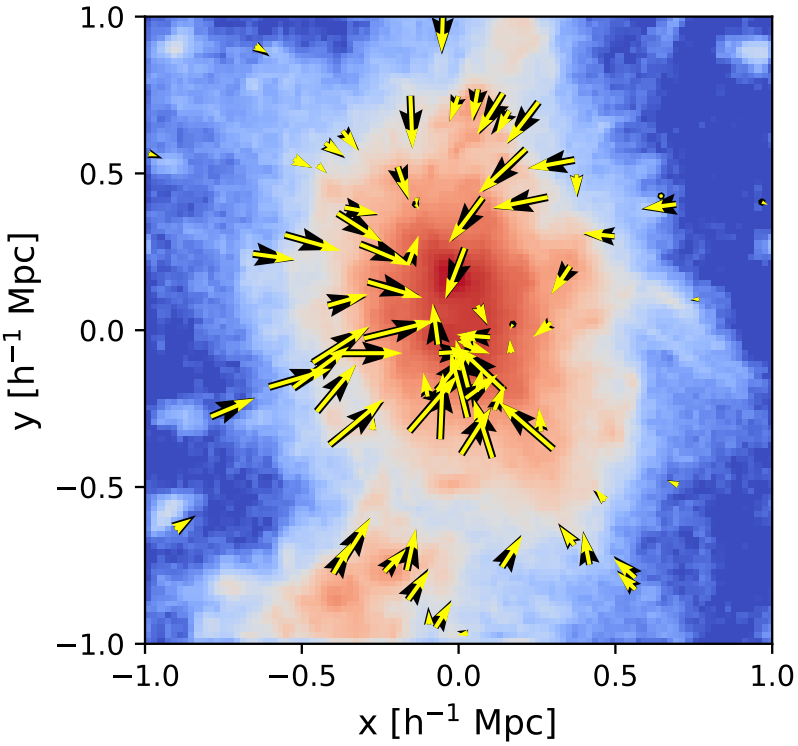



Figure 2.2: A visualization of the descent along the gradient of the gravitational potential. The thick black arrows show the displacements calculated from particle-mesh interaction (Equation 2.8), while the thin yellow color presents the displacements from particle-particle interaction (Equation 2.2). The yellow arrows mostly overlap with the black ones. Here we only show the displacements of a sample of particles in the halo.

pure dark matter high-resolution simulation (IllustrisDark) and of the hydrodynamic simulation (Illustris).

## Descent along the gradient of enthalpy

The effect of baryons on the power spectrum can be viewed from two aspects. On scales $k \approx [0.3, 30]\mathsf{hMpc}^{-1}$, pressure, stellar, and AGN feedback smooths the density field and reduces the power. On even smaller scales, the power is increased because of the cooling [e.g., 326, 342]. We will focus on $k < 10\mathsf{hMpc}^{-1}$, where the effect of cooling is sub-dominant. The two remaining effects are modeling pressure and feedback, both of which transfer matter from the high-density regions to the low-density regions.

Motivated by the hydro-PM (HPM) simulation [106], we define a pressure-like potential (spe-

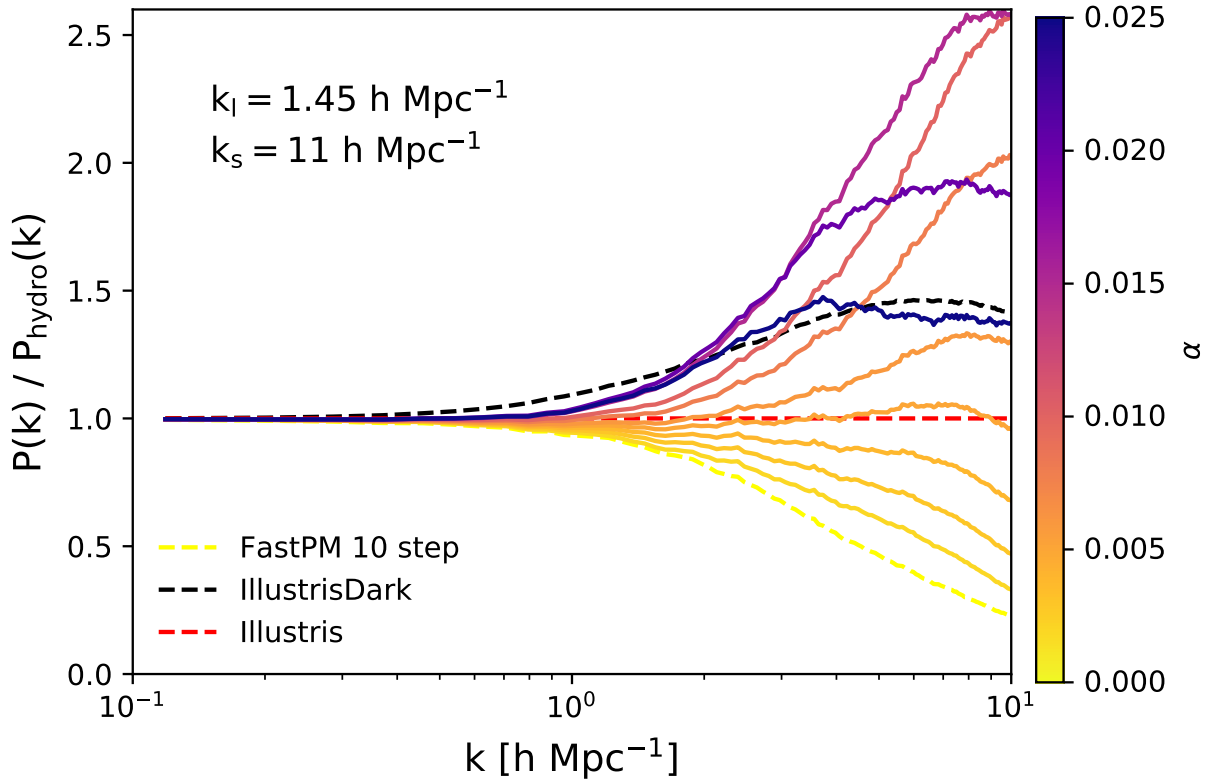



Figure 2.3: The power spectra of 10 step FastPM simulation using PGD models with different $\alpha$. At $\alpha = 0.025$ we can match the effects of pure dark matter high-resolution simulation (IllustrisDark) and at $\alpha = 0.005$ that of the hydrodynamic simulation (Illustris).

cific enthalpy). We assume that to the first order, the distribution of baryon, dark matter, and the total matter are the same. The density field is first smoothed with a Gaussian kernel

$$\hat{\mathbf{O}}_{\mathbf{J}}(k) = \exp(-\frac{(kr_{\mathrm{J}})^2}{2}) \tag{2.9}$$

where $r_J$ is the smoothing scale and we set it to be $0.1\mathrm{h}^{-1}$Mpc, of the order of the Jeans scale. We also assume a power law equation of state, [e.g. 106],

$$T(\delta) = T_0(1+\delta_b)^{\gamma-1} \tag{2.10}$$

where $T_0$ is a constant and we set it to be the characteristic temperature of IGM ($10^4$K). The HPM method typically takes $\gamma = 1.4 \sim 1.6$ for the low-density IGM. In our case, the equation of state is an effective one due to the feedback of star formation and AGN feedback, and therefore we expect $\gamma$ to be a free parameter. The pressure $P$ can be easily calculated once the temperature and density

are known:

$$P(\delta) = n_b k_B T(\delta) \tag{2.11}$$

$$= \frac{\bar{\rho_b} k_B T_0}{\mu}(1+\delta)^\gamma \tag{2.12}$$

where $\bar{\rho_b}$ is the averaged baryon density, $k_B$ is the Boltzmann constant, and $\mu$ is the averaged atomic mass of the gas which we set to be the Hydrogen atomic mass. Now we introduce the specific enthalpy $\mathcal{H}$:

$$\mathcal{H} = \frac{P(\rho)}{\rho} + \int_1^\rho \frac{P(\rho')}{\rho'}\frac{d\rho'}{\rho'} \tag{2.13}$$

The displacement is given by:

$$\mathbf{S_b} = -\frac{\beta}{H_0^2}\nabla\mathcal{H}(\delta) \tag{2.14}$$

$$= -\frac{\beta}{H_0^2}\frac{k_B T_0}{\mu}\frac{\gamma}{\gamma-1}\nabla[\hat{\mathbf{O}}_\mathbf{J}(1+\delta)]^{\gamma-1} \tag{2.15}$$

where $\beta$ is the scale factor and a free parameter in the model, $H_0$ is the Hubble parameter to make $\beta$ dimensionless. $T_0$ and $\mu$ is degenerate with the parameter $\beta$, so here we only assign them with the correct order of magnitudes and do not attempt to model their accurate values. It is clear in the above equation that the effective equation of state $\gamma$ essentially determines how the displacement depends on the density field, so we expect that changing this parameter will be able to model the halo mass dependence of AGN feedback. We note that [106] applied a pressure uniformly to all particles, while in our model each particle has the probability of $\frac{\Omega_b}{\Omega_m}$ to be identified as a baryon and hence be displaced. The rest of the dark matter particles are not displaced in this model. We found in practice that this scheme performs best for our applications. Specifically, the cross-correlation between N-body simulation and hydrodynamical simulation is higher if we assign only a fraction of particles as baryons, compared to treating all particles similarly. We will refer to this model as Enthalpy Gradient Descent (EGD) in the rest of this chapter.

## Radial flows towards the halo center

Here we describe an alternative approach that is not gradient-based. The models presented above move particles without any knowledge of where the halos are located or which halo a given particle belongs to. Instead, we asserted that the centers of halos have minimum potential and maximum density (therefore maximum pressure), even if they are not sufficiently prominent to be identified by any halo finder. If a particle is located within a halo, it is likely to be moving along the radial direction; and if it is outside the halo, the gradient will be small and so is the displacement. Alternatively, we can directly solve for the radial displacement of particles toward halo centers such that the spherical averaged radial density profile matches the halos found in hydrodynamical simulations.

In such a scheme, first, the friends-of-friends (FOF) halo finder is performed to find all the halos that we wish to calibrate the profiles in both the reference and the quasi-N-body simulations. Then the centers of halos (defined as the potential minimum)[1] are found, and particles are assigned labels according to their host halo. We assign unlabeled particles to the nearest halo.

The halo profiles and baryonic feedback are halo mass-dependent. We can divide the halos into different mass bins and measure the averaged enclosed mass as a function of radius $M_i(r)$, where $i$ denotes different halo mass bins. We represent the halo number in each mass bin as $N_i$. For the reference simulation, we sort the halos by their mass and then also divide them into different bins so that each bin has the same number of halos as $N_i$ (Abundance Matching). We measure the corresponding enclosed mass $M_{\mathsf{ref},i}(r)$ in the reference simulation. $M_{\mathsf{ref},i}(r)$ is a monotonic increasing function so we can write its inverse function

$$r_{\mathsf{ref},i}(M) = M^{-1}_{\mathsf{ref},i}(r) \tag{2.16}$$

We define the radial displacement $D(r)$ as a function of radius:

$$D(r) = r_{\mathsf{ref},i}[M_i(r)] - r \tag{2.17}$$

The definition above ensures the enclosed mass after calibration is the same with the reference simulation, so that the halo profile, which corresponds to the derivative of $M_i(r)$, also matches. Equation 2.17 also makes sure that the spherical shells of a halo do not cross with each other during the calibration.

A naive displacement may leave gaps at the edge of halos. We therefore apply a smooth truncation to the displacement according to a characteristic radius $r_{500} = (\frac{3M_{fof}}{2000\pi})$. If a particle is located outside this radius, we suppress the displacement by a factor of $\exp(1 - \frac{r}{r_{500}})$.

The full formula for the displacement is

$$\mathbf{S} = \begin{cases} D(r)\mathbf{e_r} & r \leq r_{500} \\ D(r)\exp(1-\frac{r}{r_{500}})\mathbf{e_r} & r > r_{500} \end{cases} \tag{2.18}$$

where $\mathbf{e_r}$ is the unit vector of the radial direction.

This method is the most direct way to manipulate the halo profiles, but it is not gradient-based, or based on any other physical considerations. Since this model requires lots of computations including running FOF halo finder, finding the potential minimum as the halo centers, and measuring the halo profiles, it is much slower than the potential and enthalpy gradient descent models introduced above. Reconstructing the initial conditions [345, 287] requires taking a gradient of the final density with respect to initial modes (the Jacobian), which is also problematic here, as the process involves non-differentiable procedures, such as peak finding, binning, and connecting friends-of-friends halos. The method does not enhance the internal substructures of a halo because the radial displacement is not aware of any substructures. Because of all these reasons this method is less suitable for our purposes, but we have nevertheless developed it and present results below. This method will be referred to as the Radial Flow (RF) model.

[1] Density maximum is noisy due to the small scale fluctuations.

## 2.3 Example application: matching a set of hydrodynamical and N-body simulations

As an example, we calibrate 3 "inaccurate" simulations against Illustris hydrodynamical simulations [231, 342, 341, 102]: FastPM with 10 steps, FastPM with 40 steps, and high resolution dark-matter-only simulation (IllustrisDark). These simulations use the same linear power spectrum and the same random seed as the hydrodynamical simulation. The Python version of FastPM was improved to perform the FastPM simulations used in this work [88]. The power spectra of these simulations at redshift 0 before calibration are shown in Figure 2.5:

- The 10 step FastPM lacks small scale power. Therefore we apply the potential gradient descent model to sharpen the halos and increase the power on small scales.
- The high-resolution N-body simulations are over-clustered at a small scale due to the lack of feedback processes. Therefore we mimic the baryonic effects and lower the power using our enthalpy gradient descent model.
- FastPM with 40 steps: on scales $0.5\mathsf{hMpc}^{-1} \lesssim k \lesssim 5\mathsf{hMpc}^{-1}$ it has more power than the reference simulation because of the absence of baryonic feedback, while on scales $k \gtrsim 5\mathsf{hMpc}^{-1}$, the power is reduced and is similar to 10 steps. Therefore, we apply both potential and enthalpy models to 40 steps FastPM.

We vary the free parameters in the models to fit the power spectra and find the best-fit solutions with the maximum likelihood defined in Equation 2.24.

There are 3 free parameters in the potential gradient descent model: $\alpha_{\mathsf{PM}}$, $k_l$ and $k_s$, and 2 parameters in the enthalpy gradient descent model: $\gamma$ and $\beta$. We ask $\gamma > 1$ during our fitting, so that the matter is moving from high-density regions to low-density regions.

### Visual Inspection

Before quantitatively presenting the results of our calibration, we first show a visual impression of how our models modify the matter distribution in a single halo in Figure 2.4. Compared to dark-matter-only simulation, AGN feedback in hydrodynamical simulation moves a large amount of gas from the center of halos to large radii [79, 214]. We can see this effect in Figure 2.4, where the halo from Illustris appears fuzzier at the outskirts than IllustrisDark (the upper right corner, in the upper left panel and upper middle panel). Our enthalpy gradient descent procedure successfully models this effect by pushing particles away from the halo, producing a smoother density field. The projected density field (especially at large radii) looks closer to the hydrodynamical simulation after applying our model.

Figure 2.4 also shows that the inner profile of the halo in 10-step FastPM is not cuspy enough. As discussed above, this is mostly due to the limited force resolution of particle mesh and the insufficient number of steps which limits the nonlinear effects. Both the potential gradient descent model and radial flow model produce a cuspy center by moving the particles towards the center

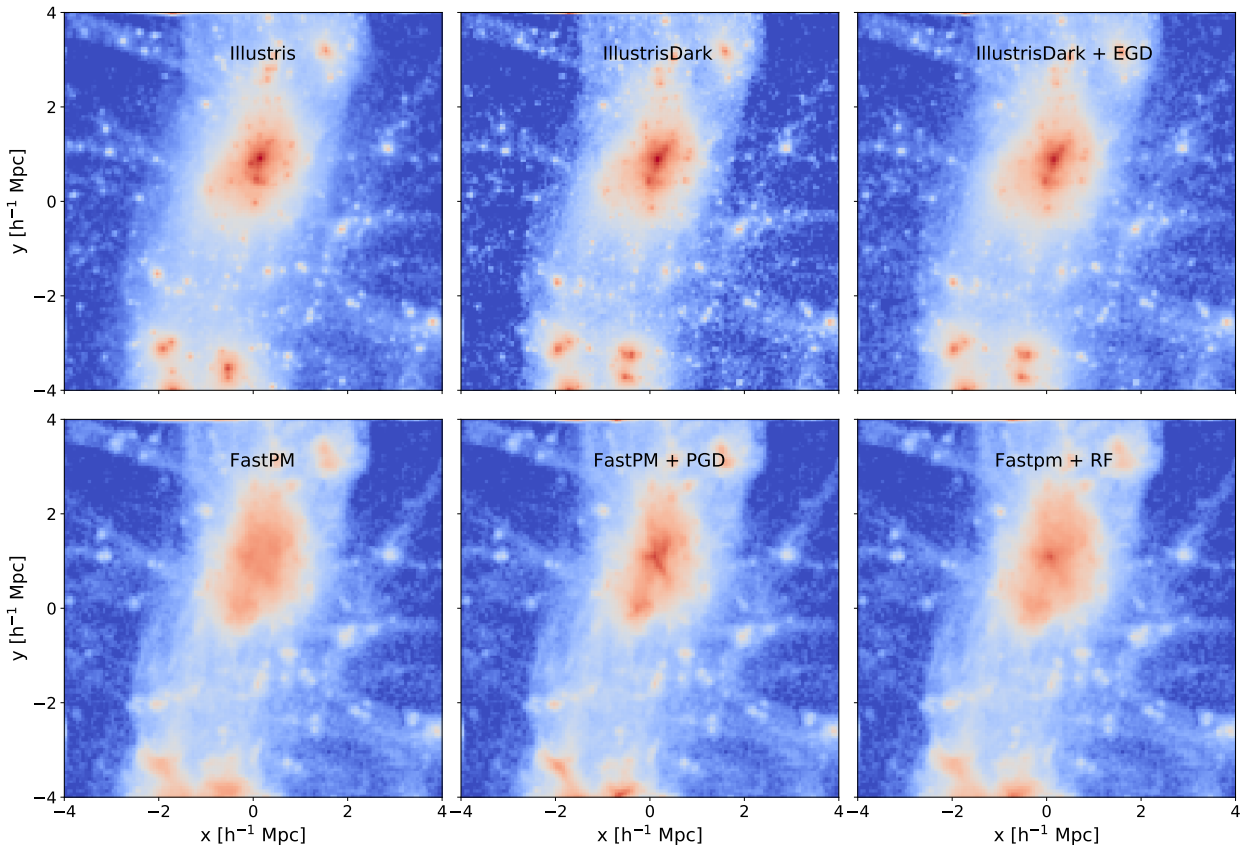



Figure 2.4: The projection of the same halo in Illustris, IllustrisDark, and FastPM, before and after applying our models. The top left panel is from Illustris-3, the top middle panel shows IllustrisDark, and the top right one is IllustrisDark after applying our enthalpy gradient descent model (see Section 2.2). The enthalpy gradient descent model moves some particles from the halo center to the outer region to simulate AGN feedback, making the outskirts of IllustrisDark closer to Illustris (see the upper right corner of each plot in the upper panel). The bottom panels show the halos in FastPM simulation with 10 steps, before and after applying our potential gradient descent model and radial flow model, respectively. The halo in FastPM does not fully collapse and no density peak can be found. After applying our models, the peaks appear. The radial flow model produces a smooth density profile with the correct spherical density profile, but the potential gradient descent model creates some of the substructures that look closer to Illustris.

and contracting the halo. However, the density field is still relatively smooth after applying the RF model because the particles are moved isotropically, and no evident substructures can be found. The potential gradient descent model, on the other hand, is able to model some of the substructures. This is not surprising. Even though halos in FastPM simulation have not fully collapsed, we expect that the seeds of these substructures remain in the density field, and the gradient descent can amplify

these density fluctuations.

In the rest of this subsection, we will present in detail how these models improve the matter power spectra, halo profiles, and sub-halo statistics.

## Power Spectra of Matter

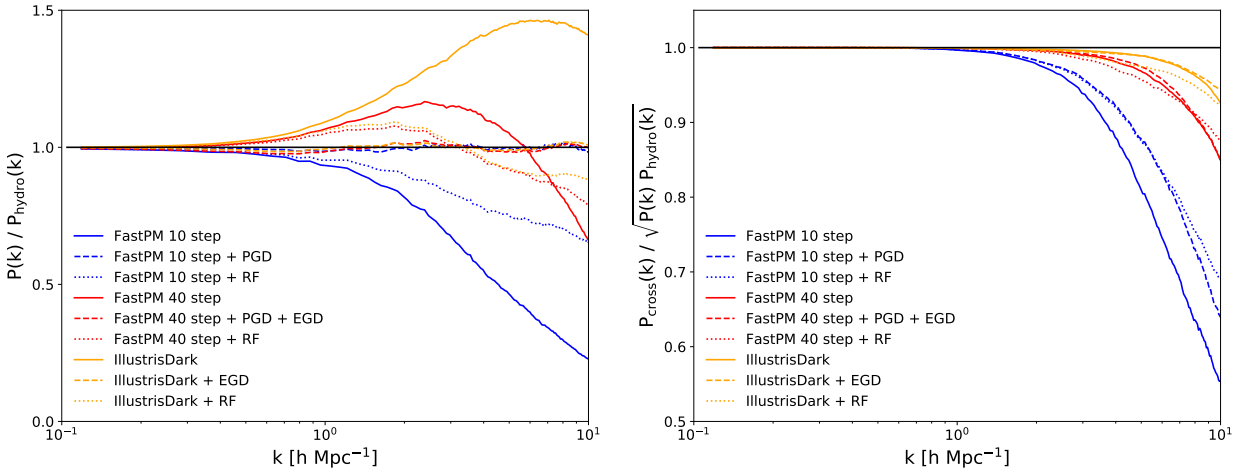



Figure 2.5: The ratio of matter power spectra (left panel) and cross-correlation coefficients (right panel) of FastPM with 10 steps, 40 steps, and dark-matter-only simulation, before and after using our models, compared to hydrodynamical simulations. The solid lines show results before calibration, while the dashed line and dotted line present the results after applying the models. After calibration, the deviations of matter power spectra compared to hydrodynamical simulation are within $5\%$. The cross-correlation coefficients also improve after calibration in most circumstances. The power spectrum is calculated using Nbodykit [118].

Figure 2.5 shows the matter power spectra and cross-correlation coefficients of FastPM and dark-matter-only simulations compared to hydrodynamical simulations, before and after applying the model. For 10 step FastPM and dark-matter-only simulation, the potential and enthalpy gradient descent models work fine, reducing the relative deviations of power spectra to within $5\%$.

We also observe that the cross-correlation coefficient of 10-step FastPM improves after calibration. We point out that the improvement is better than the 11-step FastPM simulation, although the computational cost is the same.

The RF model is not based on optimizing the power spectra, as it calibrates the halo profiles and improves the one halo term in the halo model, leading to better small-scale power spectra. We notice that the improvement of cross-correlation coefficients of PGD and RF are similar.

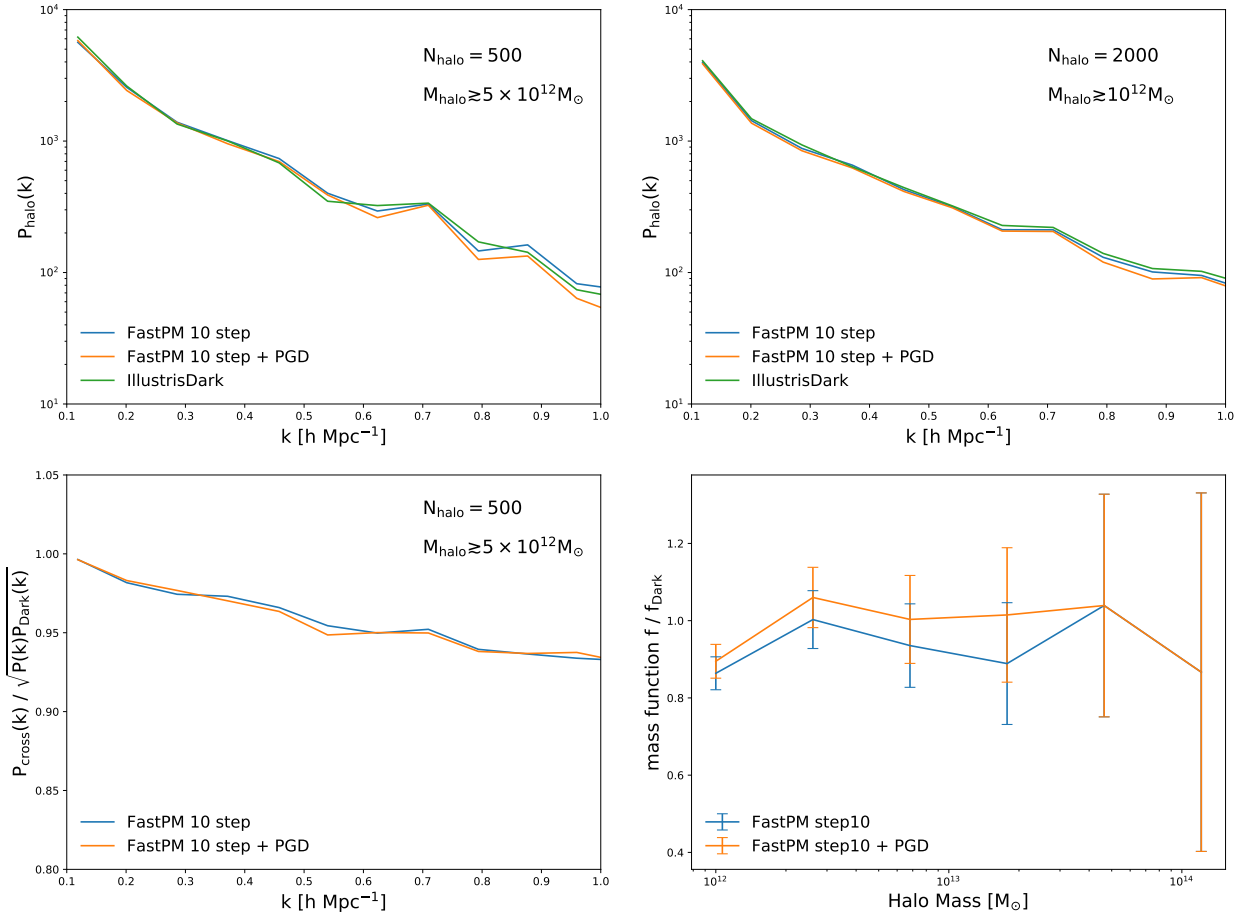



Figure 2.6: The halo power spectrum for different halo mass (upper panels), halo cross-correlation coefficient between FastPM and Illustris-Dark (lower left panel) and the ratio of halo mass function between FastPM and Illustris-Dark (lower right panel). The halo catalogs used for calculating the halo power spectrum and cross-correlation are selected from abundance matching. The upper left panel shows the power spectrum of the first 500 massive halos, which approximately correspond to halos larger than $5 \times 10^{12} M_\odot$ in Illustris-Dark; and the upper right panel shows the first 2000 massive halos, approximately corresponding to halos larger than $10^{12} M_\odot$. The power spectrum shown above has already been subtracted from the shot noise.

## Halo Statistics

Before looking into the halo inner structures, we first examine if the halo statistics are reproduced in FastPM simulation. We show the halo power spectrum, halo cross-correlation coefficient, and the halo mass function in Figure 2.6. Here we compare our results against Illustris-Dark, because we use the FOF mass to represent the halo mass and it is difficult to define an equivalent halo mass in a hydrodynamical simulation. We can see that FastPM produces similar halo statistics as the

high-resolution N-body simulation Illustris-Dark. This is consistent with the FastPM paper[89]. We also notice that the PGD model can improve the halo mass function, but has little influence on the halo power spectrum. This is as expected. The PGD model moves the particles toward the center of the halo, therefore increasing the FOF halo mass as well as the halo mass function. The centers of halos and the rank orders sorted by halo mass are almost unaffected during this process, leaving the halo power spectrum at the same abundance unchanged.

## Mass Profile of Halos

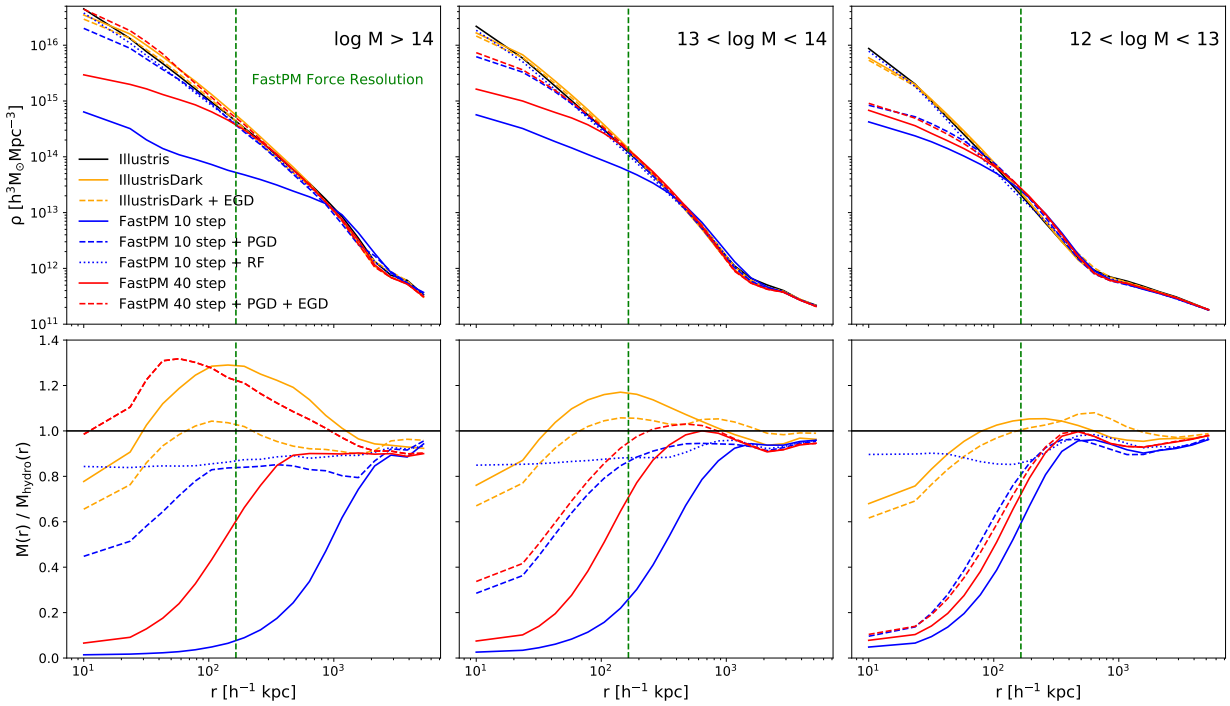



Figure 2.7: The spherical averaged halo density profile (upper panels) and enclosed mass as a function of radius (lower panels), in different halo mass bins. Black color shows Illustris, yellow represents Illustris-Dark, blue displays 10-step FastPM, and red shows 40-step FastPM. The solid line shows results before calibration, and the dashed line and dotted present results after using different calibration models. The spherically averaged matter distribution gets closer to Illustris after calibration. The effect is stronger for larger halos.

Figure 2.7 shows the calibrated FastPM and Illustris-Dark halo profile against the reference Illustris-3 simulation. We see that both density profiles and matter profiles improve after calibration. This translates to a particularly large improvement of PGD for massive halos, but for smaller halos that are barely resolved, the profile improvement is less evident. This is probably due to the fact that in smaller halos the gravitational force is too small and therefore the displacement is not enough. In

[325] it was shown that on scales $2 \lesssim k \lesssim 10h\mathrm{Mpc}^{-1}$, the power is dominated by the contribution from massive halos $m_{200} \gtrsim 10^{13.5} M_\odot$. This explains why the profiles of small halos are not very good, while at the same time, the power spectrum is matched very well.

Ideally, the halo profiles after applying the RF model should be identical to the profiles in the reference simulation, therefore the blue dotted line in Figure 2.7 should be the same as the black solid line. However, our results show differently, and it may be due to the following reasons: 1. In the RF model we define the potential minimum as the halo center (In FastPM simulation the halo profile is flatter, where the density maximum is noisy and difficult to find), but in Figure 2.7 for Illustris we choose the density maximum as the halo center. These two centers are always different in halos. 2. The halo mass increases after applying the RF model, so some halos will transit to higher halo mass bins, making average halo profiles different from the reference simulation.

### Substructures

Our numerical schemes only affect the matter distribution inside halos, leaving the clustering of halos intact. We do however expect to see an improvement in the internal structure of halos. In this section, we investigate this via the clustering of sub-halos.

The Illustris-3 simulation provides a sub-halo catalog identified by the SubFind algorithm. Here we choose the satellite sub-halos with $M > 10^{12} h^{-1}\mathrm{M}_\odot$ as our reference sub-halo catalog. For FastPM and the calibrated FastPM simulations, we use a Friends-of-Friends with a short linking length $l_{\mathrm{FOF}} = 0.05$ to find the sub-halos. Then we remove the central sub-halos, and combine those sub-halos that are closer to each other than $r < 0.2\, h^{-1}\,\mathrm{M}pc$. Finally, the rest sub-halos are abundance matched with our reference catalog.

In Figure 2.8 and 2.9 we show a visual inspection of the identified substructures before and after calibration, compared against the Illustris-3 hydrodynamical simulation. We notice that our scheme significantly increases the number of substructures in 10-step FastPM, although they usually tend to be at the wrong locations.

In Figure 2.10 we show the power spectrum, cross-correlation coefficient, stochasticity and abundance. The stochasticity is defined as the power spectrum of the residue field $\delta_{\mathrm{res}}(k)$, where $\delta_{\mathrm{res}}(k) = \delta_{\mathrm{ref}}(k) - T(k)\delta_{\mathrm{FastPM}}(k)$, and $T(k)$ is the transfer function between reference simulation and FastPM. We see that our scheme significantly improves all of the statistics. Notably, for 10-step simulation, the cross-correlation coefficient improves from 80% to 95% at the largest scale, and the stochasticity decreases to the shot noise level.

## 2.4 Parameter Selection of PGD

The selection of parameters depends on the baryonic physics models, and the parameters of the quasi-N-body simulation, such as mass resolution (typically lower than that of the hydro simulation) and number of time steps. Redshift evolution needs to be calibrated as well. On the other hand, we only have a handful of parameters, and their resolution and redshift dependencies need to be

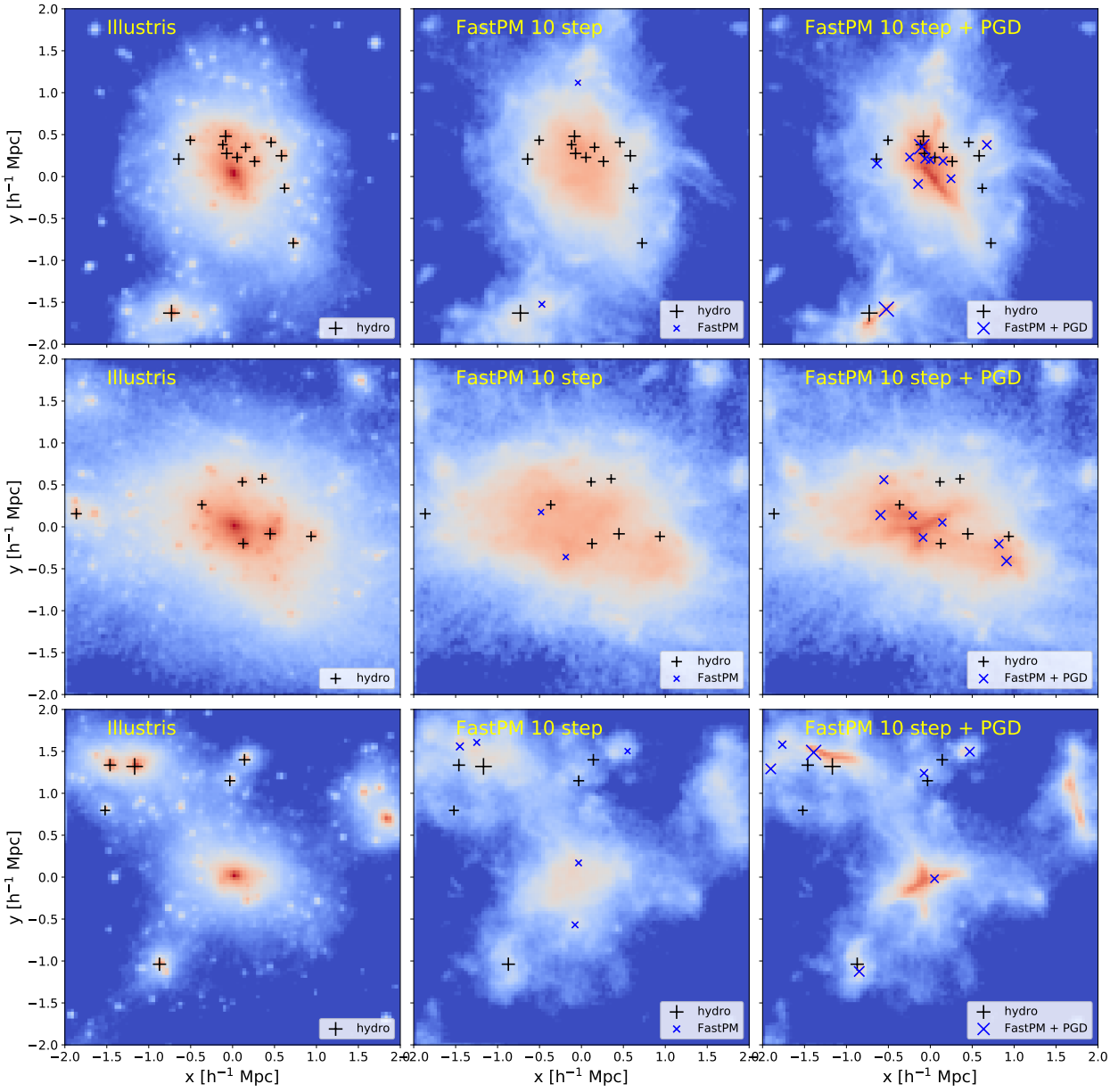



Figure 2.8: Visualizations of some large halos and their satellite sub-halos in Illustris and 10 step FastPM. The satellite sub-halos are indicated with "+" (Illustris) and "×" (FastPM). The PGD model greatly improves the identifications of substructures in 10-step FastPM.

calibrated only once. We note that for 10-step FastPM, PGD alone can provide a significant improvement on the power spectrum for a wide range of tests we tried, as seen in Figure 2.3.

We then analyzed the degeneracy of parameters with MCMC of the PGD model in Figure 2.11.

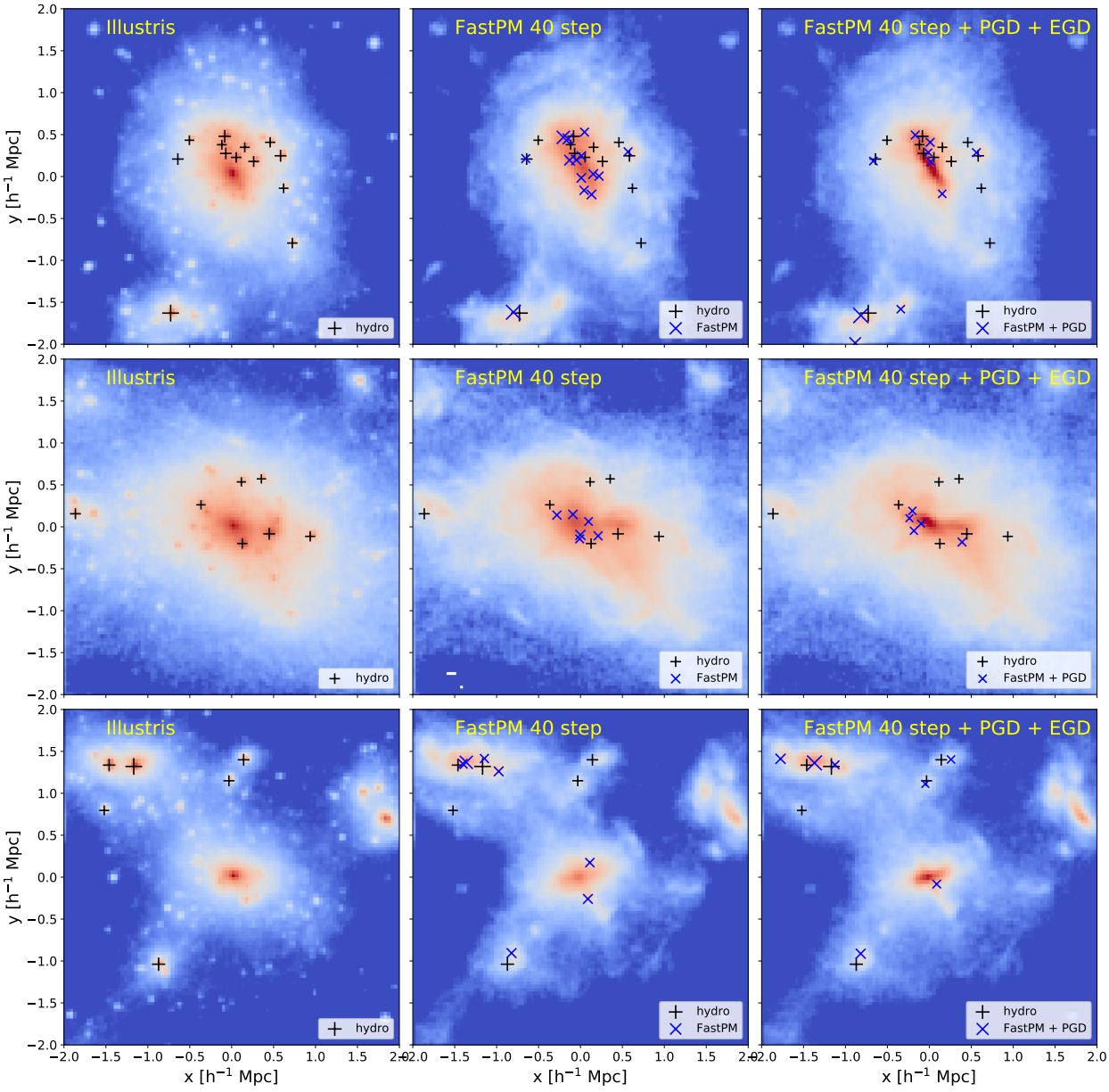



Figure 2.9: Visualizations of some large halos and their satellite sub-halos in Illustris and 40 step FastPM. The satellite sub-halos are indicated with "+" (Illustris) and "×" (FastPM).

We see that the scheme allows a wide range of $k_l$ or $k_s$ depending on the targeted simulation for calibration.

It is therefore possible to propose a relatively simple set of parameters that cover a set of resolutions. We define the dimensionless resolution parameter $\Delta = \frac{\delta}{1h^{-1}Mpc}$, where $\delta$ is the mean

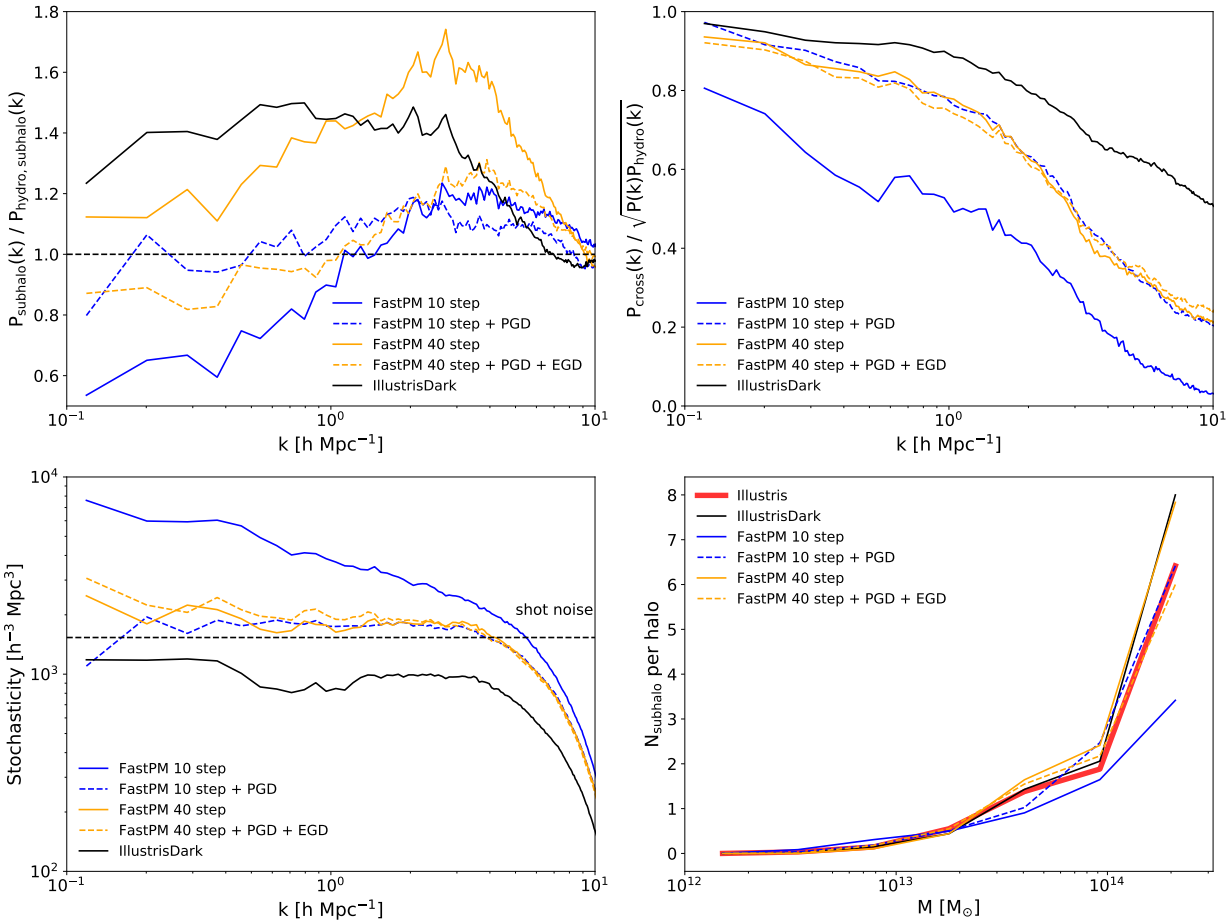



Figure 2.10: The satellite sub-halo power spectra (upper left), cross-correlation coefficients with the reference simulation (upper right), stochasticity (lower left), and the average numbers per halo (lower right). Here we show the results of FastPM 10 step (blue color) and 40 step (orange color) before (straight line) and after (dashed lines) applying our models. We also show the IllustrisDark in black color as a comparison.

separation of particles. For different resolutions and different redshifts (as cosmic scale factor $a$), we find the best-fit parameters, and then fit them with parametrized curves. This gives us an approximated empirical relation of the PGD model parameters as a function of $\Delta$ and $a$:

$$k_l = (1.52 - 0.3\Delta)\ \mathrm{h\ Mpc^{-1}} \tag{2.19}$$

$$k_s = (33.4 - 30\Delta)\ \mathrm{h\ Mpc^{-1}} \tag{2.20}$$

$$\alpha = \alpha_0 \cdot a^{\mu} \tag{2.21}$$

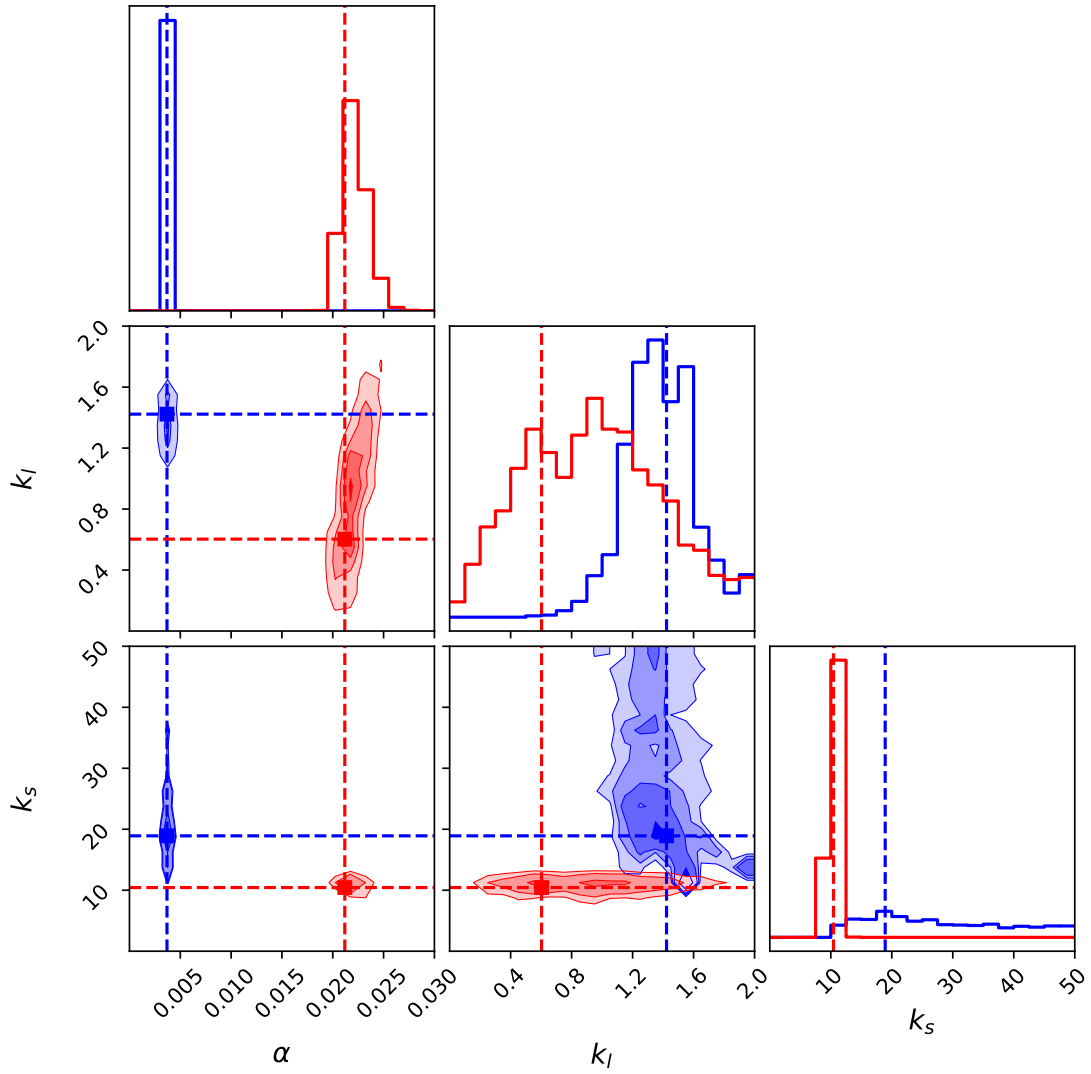



Figure 2.11: The marginalized posterior probability distribution of the 3 parameters in the potential gradient descent model when fitting the power spectra of 10 steps FastPM against Illustris (blue color) and Illustris-Dark (red color). The dashed lines show the values of the best-fit parameters. Here we set flat priors for these parameters. Because the uncertainties we choose in Section 2.7 are quite arbitrary, the contours here do not show the true locations of 1 $\sigma$ or 2 $\sigma$. However, they do show us the degeneracy of these parameters (shape of the contours) and the approximated locations of the best-fit parameters. Emcee [95] is used for sampling.

where

$$\alpha_0 = 0.0061\Delta^{25} + 0.0051\Delta^3 + 0.00314 \tag{2.22}$$

$$\mu = -5.18\Delta^3 + 11.57\Delta^2 - 8.58\Delta + 0.77 \tag{2.23}$$

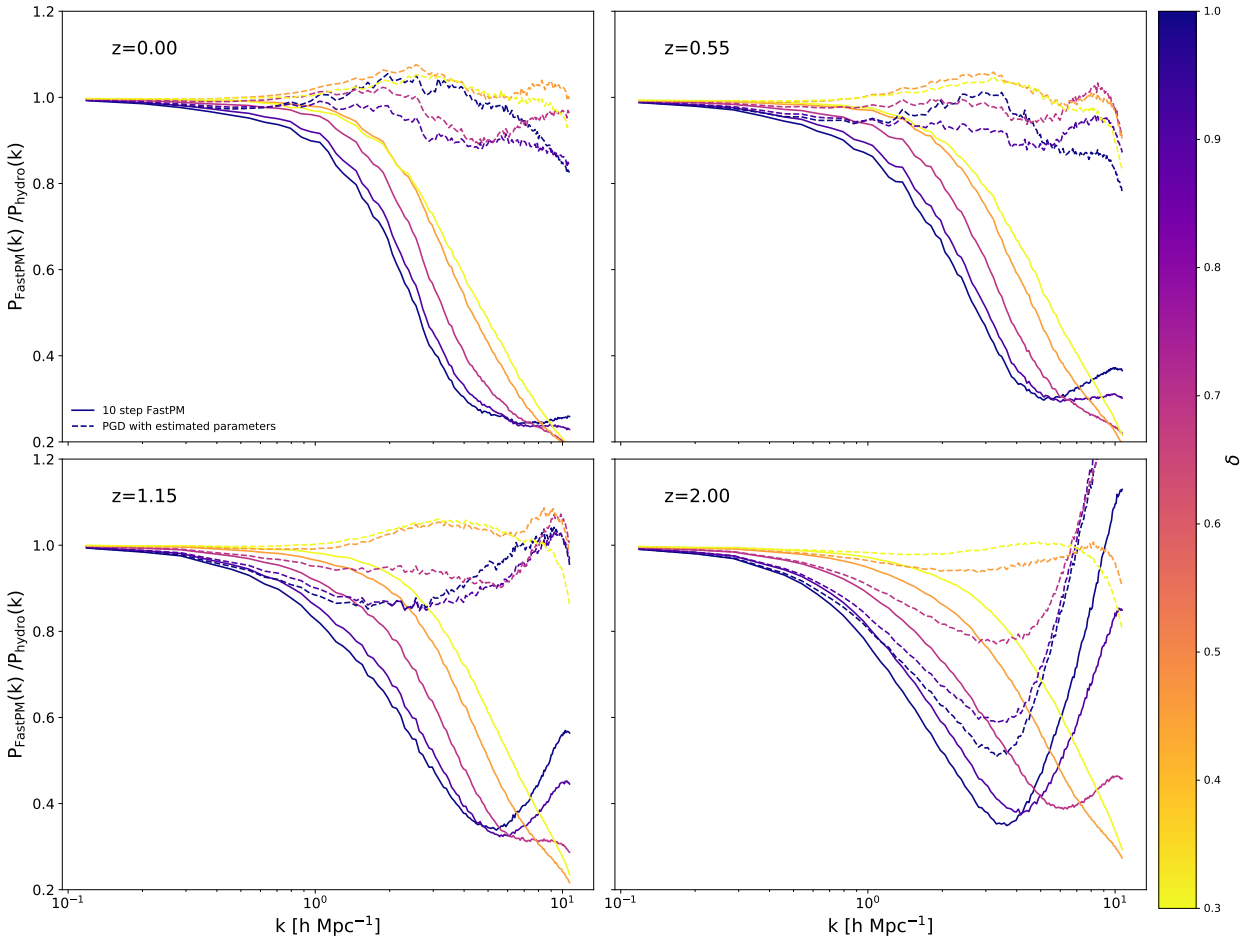



Figure 2.12: The power spectra of FastPM simulation after applying the PGD model. The parameters of the PGD model are determined using Equation 2.19 - Equation 2.23. We show the results of different redshifts and of different simulation resolutions ($\delta$ represents the mean separation of particles). At low redshifts, these parameter choices work well for all resolutions; while at high redshifts, low resolutions fail due to shot noise.

The above relation works when $0.3 < \Delta < 1$ and $0 < z < 2$.

We show the calibrated power spectra using this parameter choice in Figure 2.12. Because these relations are obtained from fitting the best-fit parameters and therefore are not accurate, the power spectrum here does not look as good as Figure 2.5. Still, they provide us a fair choice of these model parameters when we are not able to fit the power spectrum ourselves. At low redshifts, the power can be calibrated at all resolutions. At high redshift (z=2) the low resolutions do not work well, especially at small scales due to the shot noise. Therefore we recommend using high resolution ($\Delta \lesssim 0.5$) for high redshifts ($z \gtrsim 2$).

We note that the parameter choice presented above is for the Illustris baryonic model only. How-

ever, it is believed that the Illustris AGN feedback model is too strong and probably unrealistic[102, 115]. Therefore it is unknown how accurate this parameter choice is.

## 2.5 Conclusions

In this chapter, we introduce a gradient-based method to improve the modeling of matter distribution within halos in low-resolution quasi-N-body simulations. We train the method on the full N-body and hydro simulations, with the goal of making the two as close as possible in terms of summary statistics such as matter power spectrum and halo profiles. We introduce two versions of the gradient descent method. The Potential Gradient Descent model drifts the particles along the gradient of modified gravitational potential to help virialize the halos in quasi-N-body simulations. In the Enthalpy Gradient Descent model, the particles are moved along the gradient of estimated thermal pressure to model the feedback from AGN and supernovae in a hydro simulation. The latter can also be used by high-resolution pure N-body simulations to transform them into a hydro simulation. We also compare these to the Radial Flow scheme, which naively moves the halo particles along the radial directions to achieve the desired profiles. PGD and EGD are much faster and can create some of the substructures, but have very few free parameters, while RF has a lot more freedom and can work in all simulations, as long as the halo profiles are given.

We show that all of these models are able to improve the halo profiles and small-scale power spectra. The calibrations of PGD and EGD are based on fitting the power spectra. PGD sharpens the halo profiles, especially for massive halos. EGD simulates the AGN feedback by moving the matter to the outskirts of the halos. The effect of cooling and adiabatic contraction is on scales smaller than we are interested in, so here we do not attempt to model these effects. RF calibration is based on calibrating the halo profiles. Both PGD and RF improve the cross-correlation coefficients. The PGD model can also improve the subhalo statistics by magnifying the density fluctuation in the halos.

We also present empirical equations of parameter choices for 10 step FastPM, as a function of simulation resolutions and redshifts. This parameter choice gives good results at low redshifts. For high redshifts, we recommend using high simulation resolutions to reduce the shot noise and improve the small-scale power.

PGD and RF can be used to improve the dark matter field in quasi N-body simulations such as FastPM and COLA. Given the reference simulation, the parameters of PGD can be determined by optimizing the power spectrum, and this set of parameters can be used in different realizations, as long as the simulation resolutions, number of steps, and redshifts remain unchanged. We present expressions for these parameters as a function of redshifts and resolutions. These models will be particularly useful for data analyses where halo internal structures are important, such as weak lensing around galaxies and clusters. PGD will also be useful in galaxy surveys, as it improves the subhalo statistics. EGD can be used to add the baryonic effects to existing dark-matter-only simulations as a first-order approximation. One limitation of the PGD model is that it does not work that well for small halos. If halo profiles at low mass are important, one must increase the

force resolution of PM. Alternatively, one can use RF method which can give the correct average halo profile even at low halo masses.

There are three free parameters in PGD and two free parameters in EGD. These parameters depend not only on baryonic physics models, but also on simulation resolutions, redshifts, and the number of steps. These parameters cannot be derived from the first principles. To achieve the best results, they need to be optimized for different situations. This can be viewed as a positive aspect of this approach: we do not fully understand the physical processes that govern feedback models and their impact on the halo mass profiles, so these astrophysical uncertainties must be modeled with free parameters. Our approach manages to compress the number of free parameters down to a few only, so it can be used as a useful parametrization of our astrophysical ignorance that needs to be marginalized over. In this sense, one can argue that high-resolution N-body simulations are no better than low-resolution simulations: they are both missing baryonic effects, and if these effects can be incorporated with a few unknown nuisance parameters into either scheme with equal results then the advantages of the high-resolution N-body simulations are eliminated.

In the future, we plan to incorporate this scheme into FastPM and its gradient of the final density field with respect to initial density modes, which are needed for the reconstruction of initial conditions [287]. This will be particularly important if the data for reconstruction include weak lensing, which can resolve halo mass profiles, at least statistically. We also plan to investigate ways of embedding the scheme directly into the simulation as additional viscous drifting that bypasses the momentum, as a way to alleviate the redshift dependency of the parameters, and as a way to produce further enhanced weak lensing maps. Finally, it is well known that the AGN feedback model of Illustris is too strong, so in some sense the parameters we determine span the maximal range of baryonic effects. In the future we plan to test the method on several additional baryonic feedback simulations, to verify and if needed expand the parameter space of baryonic parameters, and determine their most likely values.

**Acknowledgements** We thank the anonymous referee for detailed comments that helped improve our work. The majority of the computations were performed on NERSC computing facilities Edison and Cori, billed under the cosmosim and m3058 repository. We acknowledge the support of NASA grant NNX15AL17G. We thank Simeon Bird for kindly providing the linear power spectrum and random seed of Illustris simulation.

## 2.6 Appendix A: Simulation data sets

### Illustris-3/IllustrisDark-3

Illustris [231, 342, 341, 102] is a series of cosmological hydrodynamic simulations, carried out with the moving-mesh code AREPO [299]. Each simulation evolved a periodic volume 106.5 Mpc on a side, over the redshift range $z = 127$ to the present in a $\Lambda$CDM cosmology($\Omega_M = 0.2726$, $\Omega_b = 0.0456$, $H_0 = 70.4\mathrm{km\ s^{-1}\ Mpc^{-1}}$, $n_s = 0.963$, $\sigma_8 = 0.809$). Illustris follows the evolution of the dark matter, cosmic gas, stars and supermassive black holes, with a full set of physical

models including primordial and metal-line gas cooling, star formation and evolution, gas recycling, chemical enrichment, supernova feedback and AGN feedback (for more details see [343, 320]).

Illustris has three runs (Illustris-1,2,3) at different resolutions. Since the scale we are interested in this study is larger than $k = 10hMpc^{-1}$, we focus on Illustris-3, which has a mass resolution $m_{\rm DM} = 2.8 \times 10^8 h^{-1} M_\odot$, $\bar{m}_{\rm baryon} = 5.7 \times 10^6 h^{-1} M_\odot$ and a force resolution $\epsilon_{\rm DM} = 5.68{\rm kpc}$, $\epsilon_{\rm baryon} = 2.84{\rm kpc}$. As a comparison, we also make use of Illustris-3-Dark, a dark-matter-only analog of Illustris-3.

### FastPM

FastPM is a quasi N-body particle-mesh (PM) solver, in which the drift and kick factors are modified following the Zel'dovich equation of motion so that the correct linear theory growth at large scale can be produced at a limited number of steps. We generate the initial condition with the same random seed and linear power spectrum as Illustris, starting at $z = 9$ and using the second-order Lagrangian perturbation theory. However, for simulations with a small volume ($75{\rm h}^{-1}{\rm Mpc}$) the contribution to the large-scale growth due to non-linear evolution at the box scale becomes significant (percent level), which is why the ratio of large scale power in Figure 2.5 and 2.12 is a little smaller than 1. In this chapter all FastPM simulations have the same resolution as Illustris-3 ($N_{\rm particle} = 455^3$).

## 2.7 Appendix B: Cost Function and Choice Covariance

The calibrations of potential gradient descent and enthalpy gradient descent models are based on minimizing the discrepancies of the power spectra. The covariance matrix of the power spectrum in the fully nonlinear scale is unknown. To avoid this complication, we assume a simple Gaussian likelihood that weights different scales equally.

$$p(P_{\rm ref}(k)|\theta) = \prod_{k<10{\rm h\ Mpc^{-1}}} \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp[\frac{(P_{\rm ref}(k) - P_{\rm calib}(k))^2}{2\sigma_k^2}]. \tag{2.24}$$

where $\theta$ represents the parameters in our models, and $\sigma_k$ is the error of $P_{\rm ref}(k)$. One natural choice for $\sigma_k$ is $\sigma_k = \sqrt{\frac{2}{N}} P_{\rm ref}(k)$, where $N$ is the number of k modes in this bin. However, in practice, we find that this choice often overemphasizes the small-scale power, as on small scales $\sigma_k$ is quite small due to the large number of k modes. As a result, on scale $k \approx 3{\rm h\ Mpc^{-1}}$ the fitting is quite poor, even though on smaller scales the power matches well. We argue that in observation the error of small-scale power is often dominated by systematic error, and is much larger than $\sqrt{\frac{2}{N}} P_{\rm ref}(k)$. We try to avoid this by choosing $\sigma_k = 0.1 P_{\rm ref}(k)$. The factor $0.1$ here does not change the best-fit parameters. We find that this $\sigma_k$ choice works well, as shown in the chapter.

We also attempt to maximize the cross-correlation coefficient, which gives a similar improvement in the correlation coefficient but a drastically different small-scale power spectrum.

A third cost function we attempted is directly matching the density field in configuration space by minimizing the residual,

$$y = \sum_{x_i} (\delta_{\mathsf{calib}}(x_i) - \delta_{\mathsf{ref}}(x_i))^2. \tag{2.25}$$

We find that the configuration residual down-weights the large-scale power (due to fewer modes), preferring parameters that produce incorrect large-scale power.

# Chapter 3

# High mass and halo resolution from fast low-resolution simulations

The contents of this chapter were originally published in [58],

High mass and halo resolution from fast low resolution simulations
*Dai B., Feng Y., Seljak U., Singh S.*(arXiv:1908.05276) JCAP 04 (2020) 002

Generating mocks for future sky surveys requires large volumes and high resolutions, which is computationally expensive even for fast simulations. In this chapter, we try to develop numerical schemes to calibrate various halo and matter statistics in fast low-resolution simulations compared to high-resolution N-body and hydrodynamic simulations. For the halos, we improve the initial condition resolution and develop a halo finder "relaxed-FoF", where we allow different linking lengths for different halo mass and velocity dispersions. We show that our relaxed-FoF halo finder improves the common statistics, such as halo bias, halo mass function, halo auto power spectrum, cross-correlation coefficient with the reference halo catalog, and halo-matter cross power spectrum. We also calibrate small-scale velocities of small halos to improve the power spectrum in redshift space. For matter statistics, we incorporate the potential gradient descent (PGD) method into fast simulations to improve the matter distribution at nonlinear scales. By building a lightcone output, we show that the PGD method significantly improves the weak lensing convergence tomographic power spectrum. With these improvements FastPM is comparable to the high-resolution full N-body simulation of the same mass resolution, with two orders of magnitude fewer time steps. These techniques can be used to improve the halo and matter statistics of FastPM simulations for mock catalogs of future surveys such as DESI and LSST.

## 3.1 Introduction

Numerical simulations of large-scale structure formation are essential for extracting cosmological information from current and future sky surveys. N-body simulations with semi-analytic galaxy formation models have achieved great success in cosmological analysis [165, 152, 199], but they are

also computationally expensive. Quasi-N-body PM simulations with a small number of steps such as FastPM [89] and COLA [312] provide an alternative and fast way to model galaxy statistics. It has been shown that these fast simulations predict accurate halo statistics compared to full N-body simulations of the same resolution [89, 312]. However, to generate accurate mocks for future sky surveys such as DESI [181] and LSST [195], high mass resolution and large box volumes are needed, which makes the computational cost quite high even for fast simulations. For example, DESI aims at measuring the bright emission line galaxies up to $z = 1.7$, the analysis of which requires accurate modeling of $10^{11}h^{-1}M_{\odot}$ halos [66]. Considering that using halos with less than 200 particles could lead to large systematic errors [62], and that to cover the sky up to $z = 1.7$ the box should be around $3\,h^{-1}\,\mathrm{G}pc$ per side, we need at least 4 trillion dark matter particles in the simulation. This is computationally expensive in itself even with fast simulations like FastPM, not to mention that we may need lots of different realizations to measure the covariance matrices or to study the influence of cosmological parameters. Therefore, we need to find a model that reduces the computation cost while maintaining the accuracy.

Another difficulty in these quasi-N-body simulations is the deficiency of their matter power on small scales due to insufficient force resolution. The potential gradient descent (PGD) model has been proposed to improve the modeling of matter distribution on nonlinear scales [53]. PGD was used as a post-processing correction on the static snapshot. In this chapter, we incorporate PGD into FastPM at each time step, so that it can be used in generating time-continuous light-cone mocks for weak lensing analysis.

The goal of this paper is to produce reliable predictions for halo and dark matter statistics in low-resolution FastPM simulations by training them on high-resolution N-body simulations. The plan of the chapter is as follows. In section 3.2 we try to improve the identification of small halos by modifying the FoF halo finder and removing fake halos. The small-scale velocities of halos are calibrated to improve the modeling of redshift space distortion. For the matter field, we incorporate PGD into FastPM simulation in section 3.3. By building a light-cone output we show that the method can improve the weak lensing convergence field. Finally, we conclude in Section 3.4.

## 3.2 Halo statistics and clustering

In this section, we examine and improve the halo statistics in FastPM simulation. We use IllustrisTNG [300] as our reference simulation. IllustrisTNG is a suite of cosmological hydrodynamic simulations with different box sizes and resolutions. We will mainly compare our results with TNG300-2-Dark, a dark-matter-only run in a $205\,h^{-1}\,\mathrm{M}pc$ periodic box and with $1250^3$ particles. Since previous study shows that halo statistics have around $2\%$ deviations for halos consisting of 200 particles [62], TNG300-2-Dark may not be accurate enough for the halo mass we consider (halos of 180 particles), so we also examine the TNG300-1-Dark simulation, which has an 8 times higher resolution. The agreement between TNG300-2-Dark and TNG300-1-Dark should give us an estimate of the accuracy of our reference simulation. Besides, we also show the results from the TNG300-2 hydrodynamic simulation to study the baryonic effects on these halo statistics. All these simulations share the same initial linear density field.

To perform a direct comparison with the reference simulation, we run FastPM in the same $205\,h^{-1}\,\mathrm{M}pc$ periodic box with the same linear density field by matching the random seed and linear power spectrum. We generate the initial condition at $z = 9$ using 2LPT, and then evolve the field to redshift 0 with 40 steps distributed uniformly on the scale factor $a$. Unlike TNG300-2-Dark with $1250^3$ dark matter particles, we run FastPM simulation with 8 times lower resolution, i.e., $625^3$ particles. In this section we will mostly focus on $M > 10^{11}M_\odot = 6.7 \times 10^{10}h^{-1}M_\odot$ halos, corresponding to halos of more than 22 particles in FastPM. For most comparisons in this chapter, the halo catalogs are generated from abundance matching. In the following analysis, we also compare our results with TNG300-3-Dark, which has the same resolution but is a full N-body simulation. As we will see, with standard linking length $l = 0.2$, most of the comparisons with TNG300-3-Dark are quite good, consistent with previous study [89], but the agreement is not so satisfying when compared to the higher resolution reference simulation. This suggests that halos with less than a hundred particles cannot be modeled well even with a full N-body simulation, and the deficiency of FastPM at this mass range is mostly a resolution issue. This provides additional motivation for our approach: by training on Illustris TNG300-2-Dark, which has a higher overall resolution (mass, time, and force), we can obtain results with FastPM that can exceed even Illustris TNG300-3-Dark despite its higher time and force resolution. We do so by modifying the standard Friends-of-Friends (FoF) algorithm to improve the situation for these small halos. All the halos in the Illustris TNG simulations are identified using the FoF algorithm with a linking length of 0.2.

## Relaxed-FoF

Large halos in FastPM can be modeled accurately, but small halos cannot be well resolved. For example, in Figure 3.1 we show the same halo in the high-resolution reference simulation, FastPM, and a full N-body simulation with the same resolution TNG300-3-Dark. We see that the lower mass resolution halo is more diffuse than the reference. The FoF algorithm with a standard 0.2 linking length cannot link all the particles, and breaks the halo into 2 or 3 smaller halos. As a result, FastPM and TNG300-3-Dark tend to underestimate the halo mass at this mass range, and therefore underestimate the mass function, as we can see in Figure 3.4. Because halos are broken into several small halos, they appear to be more clustered and produce a larger halo bias (Figure 3.3).

We try to improve this situation by increasing the linking length $l$ in the FoF halo finder as a function of halo mass. In the middle panel of Figure 3.1, we see that with a larger linking length, we can successfully link all the particles and reproduce the correct halo mass and position. However, increasing the linking length for all the particles will bias the halo mass for large halos, since we know that the standard linking length $l = 0.2$ is already good for large halos at redshift 0 [89]. Therefore, we make the linking length a function of the halo particle number, with a larger linking length for smaller halos. Since the linking length is not fixed, we call this method relaxed-FoF. As is shown in Figure 3.4, the 0.2 linking length predicts less massive halos at high redshifts, suggesting that the linking length should also be a function of redshift as well.

Another issue with these low-resolution simulations is that in the high-density regions, unbound clusters of nearby particles are linked by the FoF algorithm, therefore producing lots of fake halos.

**Algorithm 1** Relaxed-FoF algorithm

**procedure** RELAXEDFOF($x$, $N_p$, $l$, $r$) ▷ $x$ is the set of particles, $N_p$ is the halo bin in ascending order, $l$ is the corresponding linking length, and $r$ is the velocity dispersion threshold.
    **for** $i \leftarrow 1, \mathrm{len}(N_p)$ **do**
        $\mathrm{halo} \leftarrow$ FoF($x, l[i]$)
        **for** $j \leftarrow 1, N_{\mathrm{halo}}$ **do**
            **if** halo[j].Np $< Np[i]$ **and** halo[j].$V_{\mathrm{disp}} < r\ V_{\mathrm{std,disp}}(\mathrm{halo[j].mass})$ **then**
                save halo[j] in halocat
                remove particles that form halo[j] from $x$
            **end if**
        **end for**
        remove particles that do not form halos from $x$
    **end for**

    $L \leftarrow l[\mathrm{len}(N_p)]$
    **while** $x$ is not empty **do** ▷ keep reducing the linking length and save true halos & reject fake halos until no halos can be found
        $L \leftarrow 0.9L$
        $\mathrm{halo} \leftarrow$ FoF($x, L$)
        **for** $i \leftarrow 1, N_{\mathrm{halo}}$ **do**
            **if** halo[j].$V_{\mathrm{disp}} < r\ V_{\mathrm{std,disp}}(\mathrm{halo[j].mass})$ **then**
                save halo[j] in halocat
                remove particles that form halo[j] from $x$
            **end if**
        **end for**
        remove particles that do not form halos from $x$
    **end while**
**end procedure**

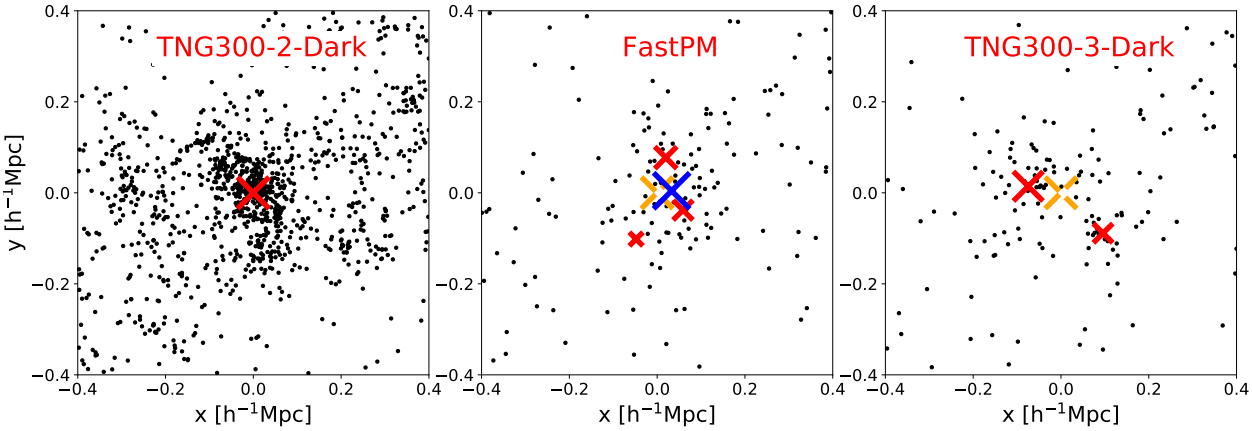



Figure 3.1: The projected map of the same halo in TNG300-2-Dark (left panel), FastPM (middle panel) and TNG300-3-Dark (right panel). The halo mass is around $1.6\times10^{11}h^{-1}M_{\odot}$, corresponding to 425 particles in TNG300-2-Dark simulation, and 53 particles in FastPM and TNG300-3-Dark. We perform the standard FoF algorithm with a linking length of 0.2 on each of them, and the halo centers of mass are represented as red crosses (the size of the cross is proportional to the halo mass). The blue cross in the middle panel shows the center of mass of the halo identified with a larger linking length. In the middle and right panels, we also show the true position (identified in TNG300-2-Dark) as orange dashed crosses.

With larger linking lengths, we expect this issue to be more severe. We find that these fake halos are likely to have larger velocity dispersion. This is expected, since the particles that make up those fake halos are "accidental" close neighbors and are not gravitationally bound. Therefore, for each small halo we calculate the quantity $r = \frac{V_{disp}}{V_{std,disp}(M,z)}$, where $V_{disp}$ is the velocity dispersion we measured from simulation, and $V_{std,disp}(M,z)$ is the expected velocity dispersion of a halo at this mass predicted by the common scaling relation [87].

$$V_{std,disp}(M,z) = V_0 \left(\frac{E(z)M}{10^{15}h^{-1}M_{\odot}}\right)^{\alpha} \tag{3.1}$$

where $V_0 \simeq 1100\mathrm{kms}^{-1}$, $E(z) = H(z)/H(0)$ is the dimensionless hubble parameter, and the slope $\alpha$ is around $0.3$. If the quantity $r$ is larger than a threshold $r_0$, we consider the halo as a fake one and reject it from the halo catalog. Since the fake halos we remove are mostly in the high-density regions, we expect this procedure to reduce the bias of small halos.

We increase the linking length for small halos to better identify the halos in low-resolution quasi-N-body simulations, and use velocity information to help remove the misidentified halos. Several previous papers have made similar attempts and achieved good results. For example, [205] generated PTHalos from a 2LPT field using a much larger linking length ($b = 0.38$), and then calibrated the halo mass to match the given halo mass function. Iteratively decreasing the linking length is commonly used in sub-halo finders (e.g., HFOF [169], ROCKSTAR [19]). Particle

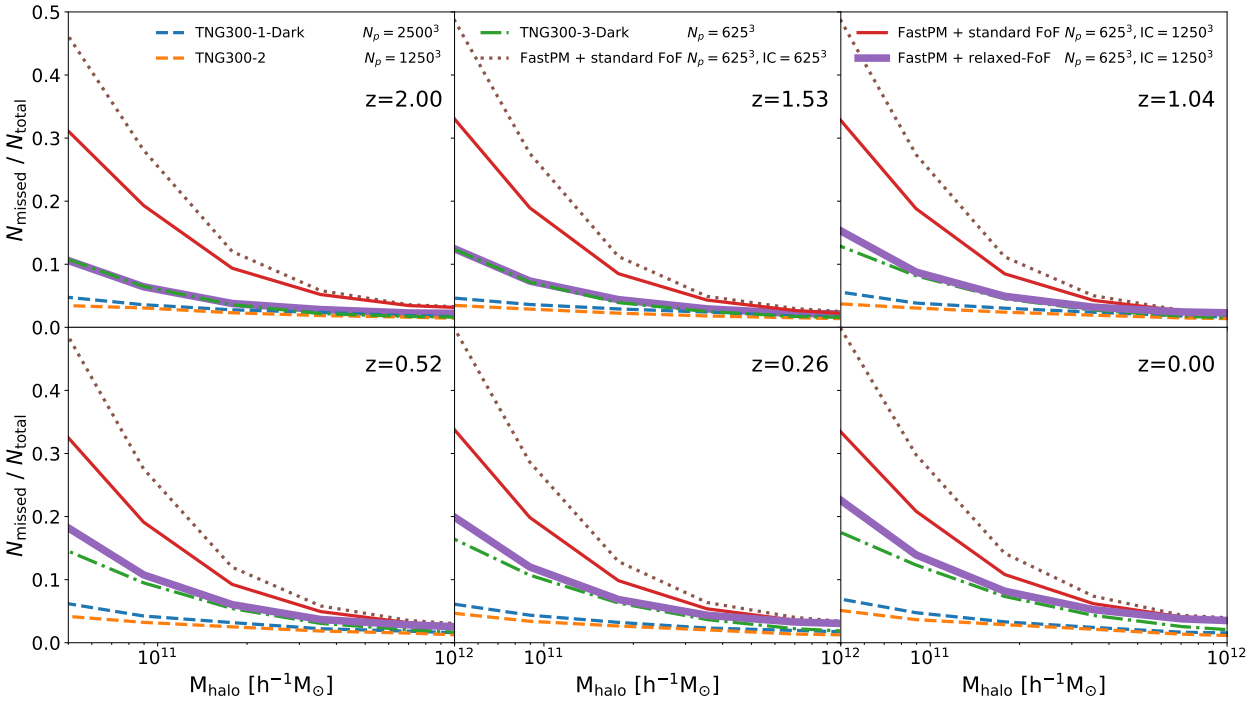



Figure 3.2: The ratio of missed halos as a function of halo particle number in different simulations and various redshifts. Here we choose TNG300-2-Dark as the reference (same for Figure 3.3 to Figure 3.8). The nonzero ratio of missed halo in the massive end is mostly because of the bridging effect.

velocities have been incorporated in many modified FOF halo finders and phase space finders(e.g., 6DFOF [68], HSF [200], ROCKSTAR [19]), and many algorithms remove particles that are not dynamically bound to the halos. Considering that the small-scale velocities are not very accurate in our simulation, here we do not try to find halos in the phase space, and only use the dynamical information to decide whether a halo is real or not. We also tried removing unbounded particles from the halos, but we did not find improvements in our case.

We divide the halos into several bins $N_{p,i}$ ($N_{p,i}$ is the maximum halo particle number of bin $i$), and for each bin we have the corresponding linking length $l(N_{p,i}, z)$. We first run the FoF halo finder on all the particles in the snapshot with linking length $l(N_{p,1}, z)$ for the smallest halo bin. Then we select all the halos that are larger than the halo particle number $N_{p,1}$ and the halos that are rejected by the velocity dispersion criterion, and rerun the FoF halo finder on the particles that form these halos with the linking length $l(N_{p,2}, z)$ of the next bin. We repeat this procedure until we finish the largest halo bin. For the rest of the particles that form the fake halos, we keep running the FoF halo finder and reducing the linking length, with fake halos rejected at each iteration, until there are no particles left. The function $l(N_{p,i}, z)$ and $r_0(z)$ are simple functions we choose to produce the correct halo mass function and halo bias:

$$N_{p,i} = \{20, 40, 80, 160, 320, \inf\} \tag{3.2}$$

$$l(N_{p,1}, z) = l_1 - \frac{A_1}{1+z} \tag{3.3}$$

$$l(N_{p,6}, z) = \max(l_6 - \frac{A_2}{1+z},\ 0.2) \tag{3.4}$$

$$l(N_{p,i}, z) = \frac{(6-i)N_{p,1} + (i-1)N_{p,6}}{5} \tag{3.5}$$

$$r_0(z) = B_1 - B_2 \log(1+z) \tag{3.6}$$

where $l_1$, $l_6$, $A_1$, $A_2$, $B_1$ and $B_2$ are free parameters. In our setup we find $l_1 = 0.25$, $l_6 = 0.235$, $A_1 = 0.012$, $A_2 = 0.06$, $B_1 = 4.28$ and $B_2 = 2.17$ give us good halo statistics for $0 \leq z \leq 2$.

Even though relaxed-FoF calls a standard FoF algorithm more than 6 times, it does not take 6 times longer, because after each iteration the number of remaining particles quickly decreases. For example, only about $50\%$ particles are left after the first iteration. In practice, we find that relaxed-FoF normally takes around twice as much time as standard FoF.

In addition to improving the halo finder algorithm, we also find that the small-scale power in the initial condition is crucial for the identification of small halos. We find it necessary to generate the linear density map with a mesh that is twice finer than the particle grid, which helps to improve the various halo statistics (Figure 3.2 to Figure 3.6). We tried further increasing the resolution of the initial condition, but the halo statistics did not improve. Note that in this study the force resolution is also twice the resolution of particles, so we can use the same force mesh to generate the initial condition, and increasing the resolution of IC does not require more memory than standard FastPM.

Before we examine any halo statistics in the next subsection, we first take a look at how well each individual halo can be reproduced. If two halos from two simulations are within $0.4\,h^{-1}\,\mathrm{M}pc$ and if their mass is within a factor of 2, then we say they are the same halo, and each halo cannot be matched with more than one halo. In Figure 3.2 we show the ratio of missed halos as a function of halo mass for different redshifts. We define missed halos as the halos that cannot find a counterpart in the other simulation. We see that as we go to smaller halos, the ratio of missed halos increases and reaches around $18\%$ ($25\%$ if the linear density map has the same resolution as particle grid) at $10^{11}h^{-1}M_\odot$ halos for FastPM with constant 0.2 linking length. After switching to our relaxed-FoF, the ratio of missed halos decreases at all redshifts and all halo masses. The improvement is larger at higher redshift. In particular, the ratio of missed halos is reduced to about $6\%$ for $10^{11}h^{-1}M_\odot$ halo at redshift 2, comparable to the full N-body simulation TNG300-3-Dark.

In Figure 3.2 we also show the ratio of missed halos for higher resolution N-body simulation TNG300-1-Dark and hydro simulation TNG300-2. Here the ratio is not zero even for large halos, due to the bridging effect. If two nearby halos are linked together in one simulation, while they are identified as two separate halos in another simulation, they will not be matched using our algorithm and therefore produce a nonzero fraction of missed halos.

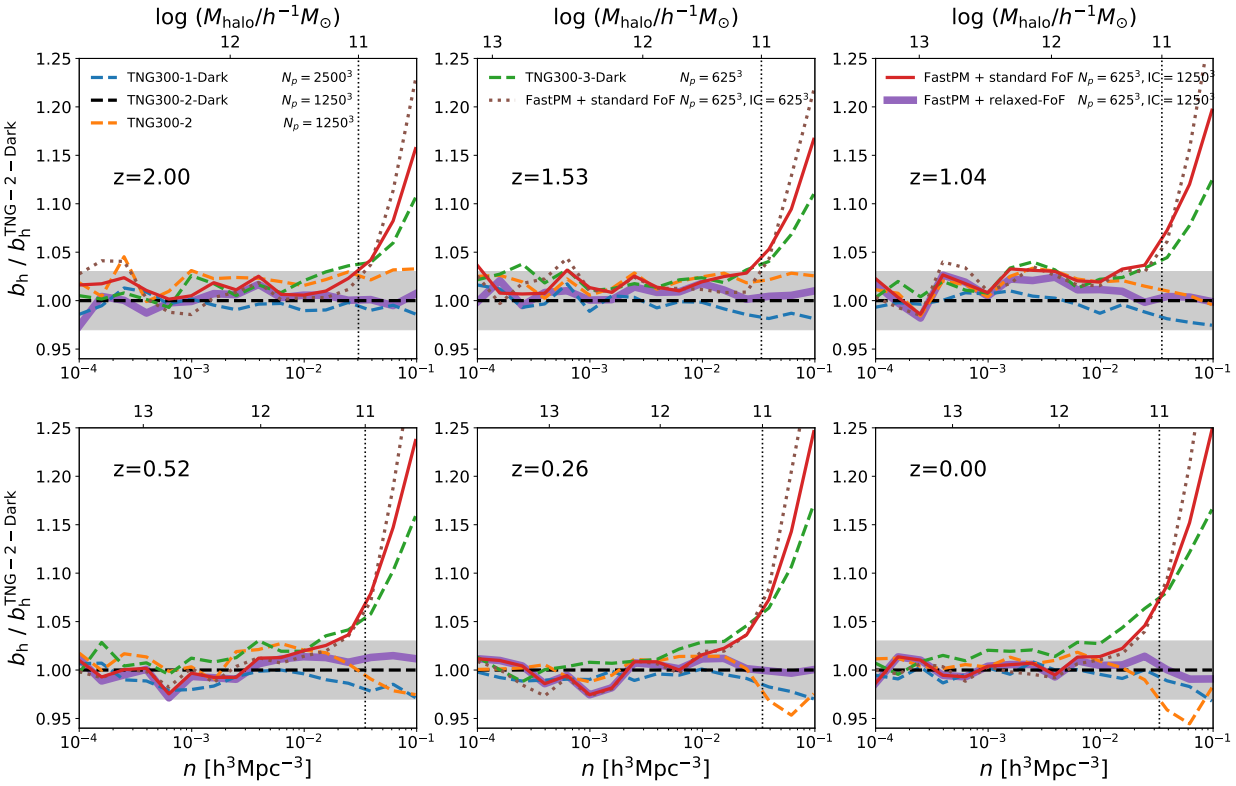



Figure 3.3: The ratio of halo bias as a function of abundance measured in different simulations and various redshifts. The halos are selected by abundance matching, and the x-axis shows the halo abundance. Larger abundance means smaller halos. The dotted vertical line shows the abundance of a $10^{11}h^{-1}M_\odot$ halo, corresponding to 33 particles. The shaded region represents $3\%$ deviation. The power spectrum is calculated using Nbodykit throughout the paper [118].

## Halo statistics in real space

We first examine the halo bias defined with the halo-matter cross-correlation

$$b = \lim_{k\to 0} \frac{P_{hm}(k)}{P_{mm}(k)}. \tag{3.7}$$

We present the halo bias results in Figure 3.3. We see that the bias given by different simulations fluctuates a lot even for the largest halos (lowest abundance). This is because the halo mass is scattered in different simulations so the same abundance does not guarantee the same halo catalog. However, comparing the three N-body simulations of different resolutions, we can see a tendency that higher resolution simulation shows a lower halo bias, especially for small halos. Similarly, FastPM also gives a very high bias for small halos, mostly due to its low resolution, but with our relaxed-FoF halo finder, the halo bias is brought down to the normal level.

With larger linking lengths, the small halos can be better identified, so the improvement of the halo mass function at the low mass end is expected (shown in Figure 3.4). With the same linking

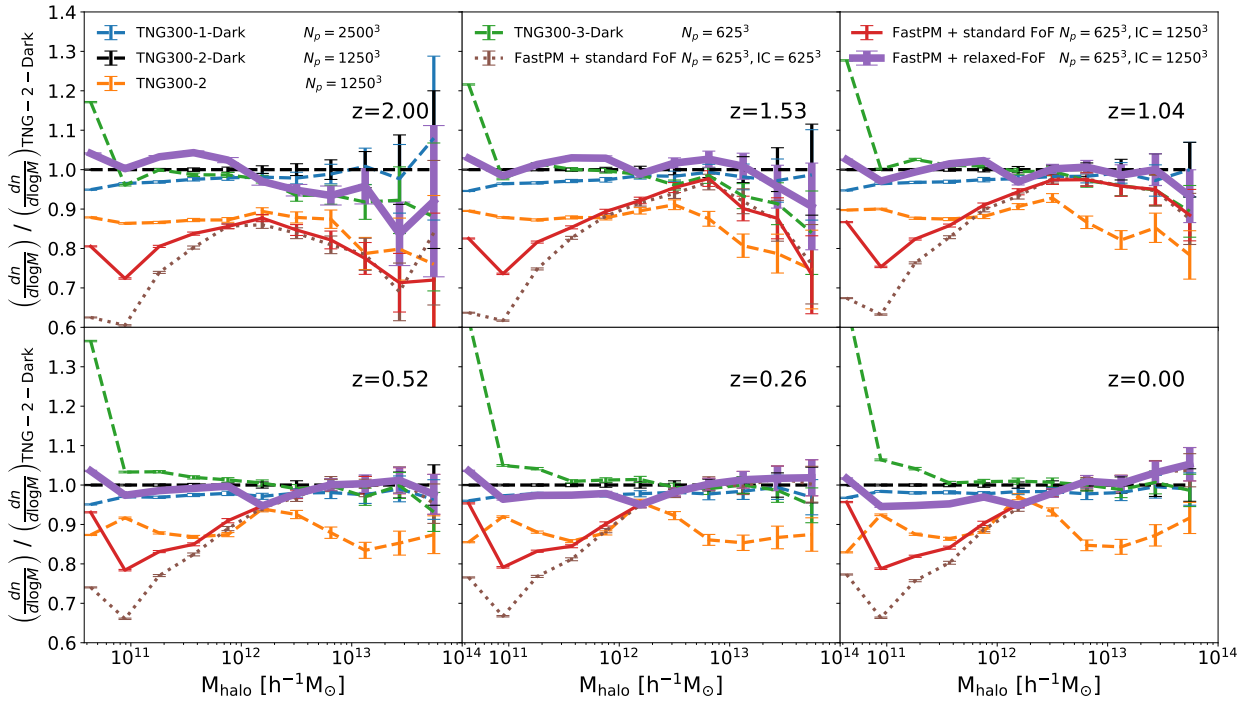



Figure 3.4: The ratio of halo mass function from different simulations and various redshifts. The halo mass here is defined as the FoF mass. For the hydro simulation, the FoF algorithm is run on the dark matter particles, and baryon particles are attached to the same groups as their nearest dark matter particle. Poisson noise errors are shown.

length 0.2, FastPM shows a large discrepancy with full N-body simulations of all resolutions, suggesting that this deficiency in halo mass function is not due to the resolution effect, but a failure of the FoF with standard linking length to resolve small halos in FastPM, which relaxed-FOF corrects for. Another interesting feature is that the baryonic feedback seems to reduce the mass function by $10\%$ to $20\%$, but this comparison is based on FoF mass and it is unclear if it is a meaningful comparison against hydrodynamic simulations.

Next, we select halo catalogs from different simulations with abundance matching, and examine their auto power spectrum (Figure 3.5), halo-matter cross power spectrum (Figure 3.6), and the cross-correlation coefficient with the reference simulation TNG300-2-Dark (Figure 3.7). The halo catalogs correspond to $M \geq 10^{11} M_\odot = 6.8 \times 10^{10} h^{-1} M_\odot$ halos (22 particles in FastPM). As mentioned above, the same abundance does not guarantee the same halos, so we expect to see a little scatter across different simulations. As long as the deviation of FastPM is comparable to the scatter of TNG-1-Dark or TNG-2, we can say the predictions of FastPM are equivalent to those of more expensive high-resolution full N-body simulations.

We see that with standard $l = 0.2$ linking length, the halo auto power spectrum and the halo-matter cross power spectrum of FastPM are very similar (slightly worse) to those of the

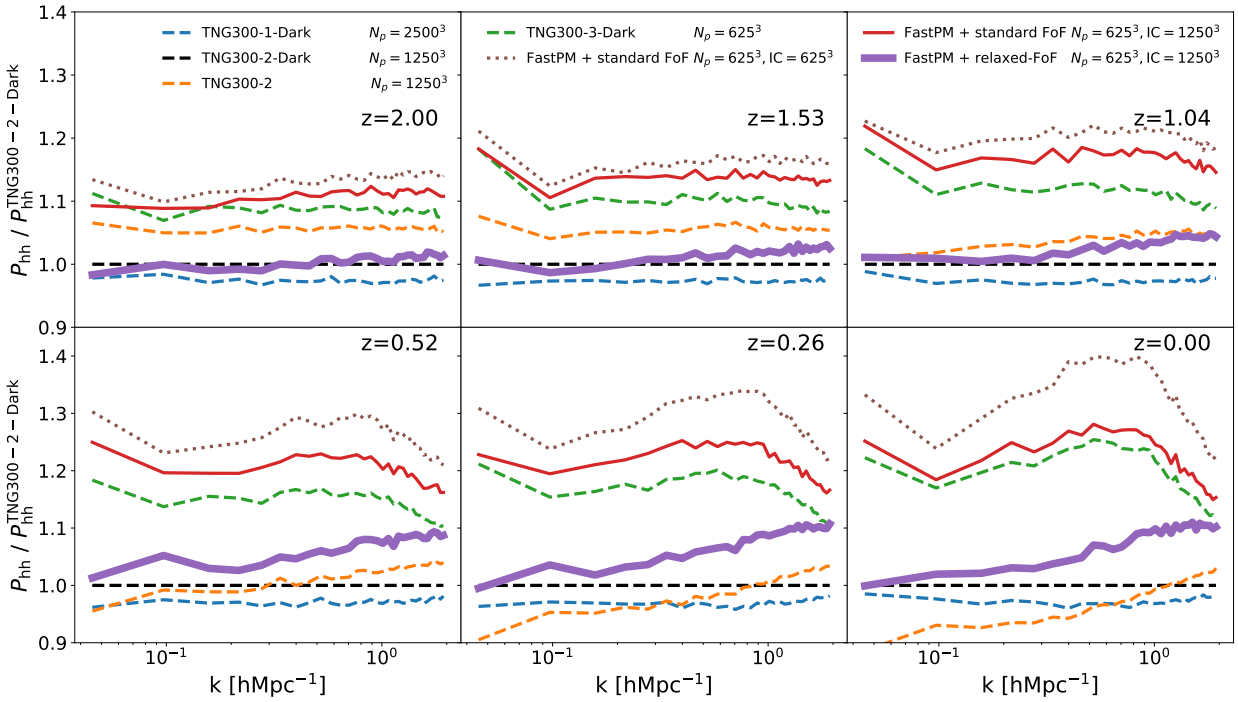



Figure 3.5: The ratio of halo auto power spectrum from different simulations and various redshifts. Similar to Figure 3.2, TNG300-2-Dark is chosen as our reference. The halos are selected using abundance matching, corresponding to $M \geq 10^{11} M_\odot = 6.8 \times 10^{10} h^{-1} M_\odot$ halos (22 particles in FastPM.)

full N-body simulation with the same resolution. After changing to relaxed-FoF, the halo auto power spectrum and halo-matter cross power spectrum improve on all scales and all redshifts. In particular, their deviations from TNG300-2-Dark are consistent with the scatter induced by abundance matching at high redshift ($z \geq 1$), while at low redshift ($z \leq 0.5$) on small scales FastPM is overpredicting power. Note that at low redshift, even though the auto power spectrum of FastPM is not consistent with our reference N-body simulation on small scales, its slope is actually quite similar to the prediction of TNG300-2 hydrodynamical simulation. The cross-correlation coefficient is consistent with the ratio of missed halos (Figure 3.2), that after improving the halo finder the halo catalog from FastPM is similar to a full N-body simulation TNG300-3-Dark.

Here in this section we only show the results of $M \geq 10^{11} M_\odot = 6.8 \times 10^{10} h^{-1} M_\odot$ halos. The auto power spectrum of larger halos is shown in Appendix 3.5.

## Velocity calibration and halo statistics in redshift space

In Figure 3.8 we show the halo auto power spectrum in redshift space. On large scales, similar to the situation in real space, relaxed-FoF improves the power. On small scales, however, the velocities of small halos cannot be accurately modeled by low resolution FastPM. Even with relaxed-FoF,

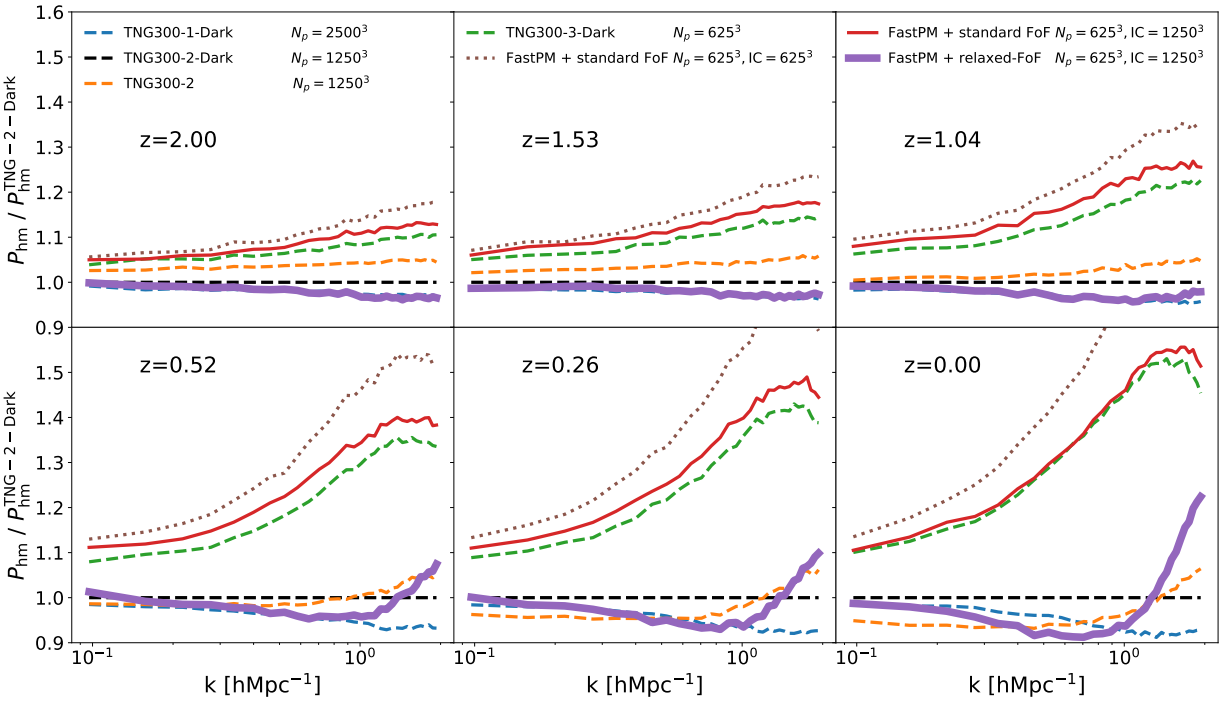



Figure 3.6: The ratio of halo-matter cross power spectrum from different simulations and various redshifts.

FastPM predicts too much power on small scales in redshift space. We try to calibrate the velocities by adding an irrotational velocity term to the particles inside small halos:

$$\mathbf{u}_{\rm calib} = \mathbf{u}_{\rm COM} - \left\langle \frac{1}{a}\nabla V \right\rangle \exp(-N/N_c) \tag{3.8}$$

where $\mathbf{u}_{\rm COM}$ is the measured halo center of mass velocity, $N$ is the number of particles in the halo, and $N_c = 160$ is the halo size which we believe can be modeled well by FastPM (shown in Appendix 3.5). We do not intend to modify the velocities of the large halos, so we put a factor $\exp(-N/N_c)$ in the equation. Since we assume the new velocity component is irrotational, it can be written as the gradient of velocity potential $V$, and we average over all the particles in the halo to give a center-of-mass velocity correction. We try to learn the velocity potential $V$ from the matter gravitational potential $\Phi$, by assuming

$$V(a,k) = T(a,k)\Phi(a,k) \tag{3.9}$$

where $a$ is the scale factor, and the transfer function $T(a,k)$ should go to zero when $k$ goes to zero to prevent any modification of large-scale velocities. We further assume

$$T(a,k) = (C_1 a - C_2)\exp\left(-k_c^2/k^2\right)\exp\left(-k^2/k_s^2\right) \tag{3.10}$$

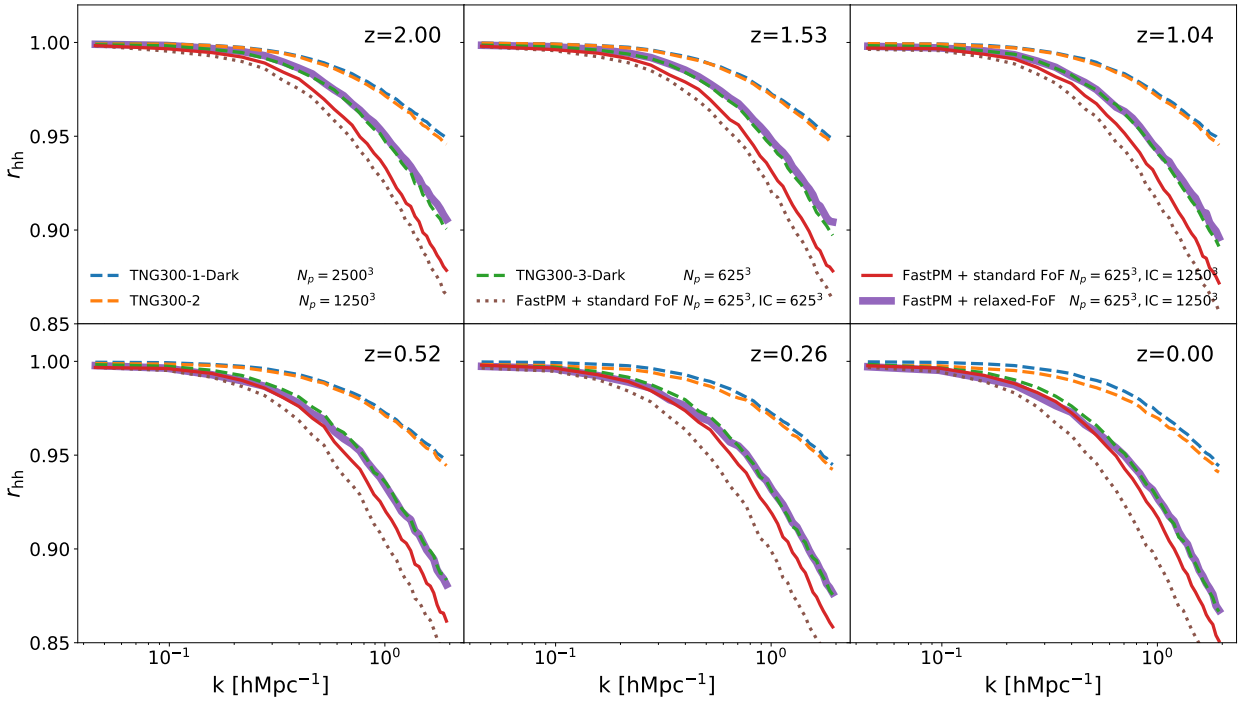



Figure 3.7: The cross-correlation coefficient of the reference halo with halos from other simulations in different redshifts.

where $k_c$, $C_1$ and $C_2$ are free parameters. We find $k_c = 1\mathrm{h}^{-1}\mathrm{Mpc}$, $C_1 = 0.23$, and $C_2 = 0.03$ give us good small-scale halo power in our setup. We introduce the factor $\exp\left(-k^2/k_s^2\right)$ to reduce the numerical effect induced by the mesh resolutions, and $k_s$ is fixed to $5\mathrm{h}^{-1}\mathrm{Mpc}$. The auto power in redshift space of halos larger than $10^{11}M_\odot = 6.8 \times 10^{10}h^{-1}M_\odot$ is shown in Figure 3.8. The results of other mass bins are shown in Appendix 3.5. Noting the similarity between Equation 3.8, 3.9, 3.10 and Equation 3.11, 3.12, our velocity calibration model can be seen as applying Potential Gradient Descent (PGD) model to velocities, where both methods model correction vectors as the gradient of modified gravitational potential.

## 3.3 Dark matter statistics

In this section, we focus on improving the matter distribution on small scales. We first incorporate the PGD model into every step of FastPM, and show that the redshift evolution of the PGD parameters can be parametrized by simple analytical functions. Then we build a light-cone simulation with the output of FastPM, and show that the PGD model can improve the weak lensing convergence power spectrum.

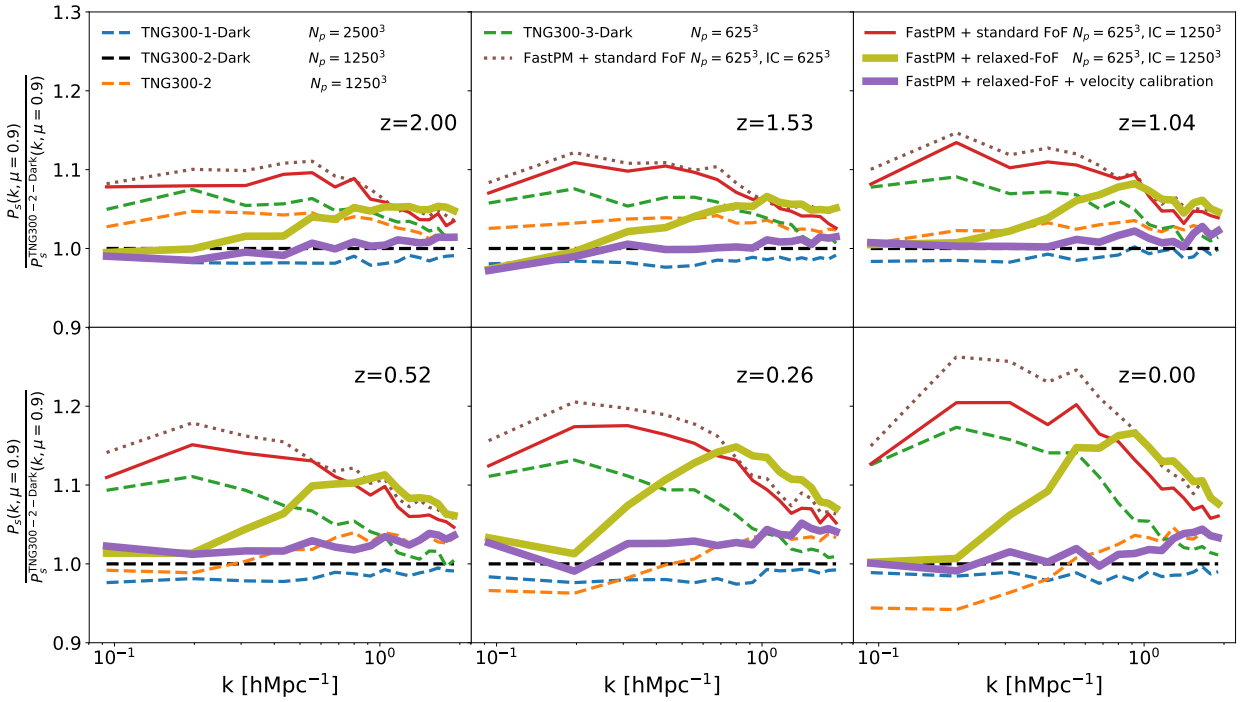



Figure 3.8: The ratio of the halo power spectrum in redshift space from different simulations and various redshifts. Here we only show the power spectrum of $k$ parallel to the line of sight, i.e., $\mu = 0.9$, where $\mu = k_{\parallel}/k$. The $k$ mode perpendicular to the line of sight is not affected by RSD, and therefore the RSD halo power spectrum with $\mu = 0$ is similar to the halo power spectrum in real space presented in Figure 3.5.

## PGD embedded in FastPM

The basic idea of the potential gradient descent (PGD) model is to add an additional displacement on the output position of particles to mimic the missing sub-grid physics in simulations. The additional displacement is modeled by the gradient of a modified gravitational potential, given by

$$\begin{aligned} \mathbf{S} &= (\alpha/H_0^2)\ \nabla(\hat{\mathbf{O}}_\mathbf{l}\hat{\mathbf{O}}_\mathbf{s}\phi) \\ &= (4\pi G\bar{\rho}\alpha/H_0^2)\ \nabla(\hat{\mathbf{O}}_\mathbf{l}\hat{\mathbf{O}}_\mathbf{s}\nabla^{-2}\delta) \end{aligned} \tag{3.11}$$

where $\phi$ is the gravitational potential field, $\delta$ is the matter overdensity, $\bar{\rho}$ is the averaged matter density, $\alpha$ is a free parameter, $\hat{\mathbf{O}}_\mathbf{l}$ and $\hat{\mathbf{O}}_\mathbf{s}$ are a high pass filter and a low pass filter, respectively

$$\hat{\mathbf{O}}_\mathbf{l}(k) = \exp\left(-\frac{k_l^2}{k^2}\right), \tag{3.12}$$

$$\hat{\mathbf{O}}_\mathbf{s}(k) = \exp\left(-\frac{k^4}{k_s^4}\right). \tag{3.13}$$

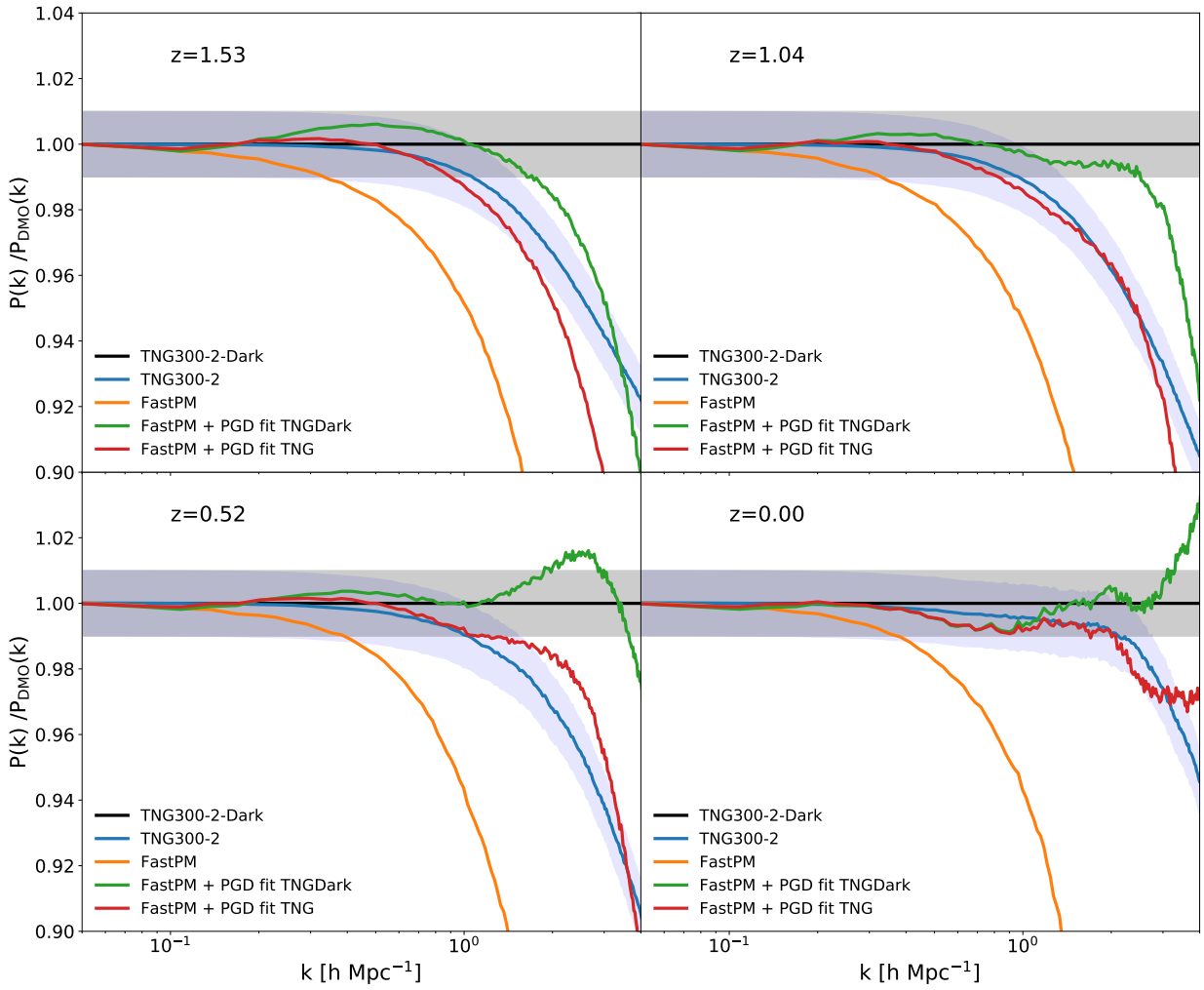



Figure 3.9: The matter power spectrum of FastPM simulation, before and after the calibration, and the reference simulation (TNG300 and TNG300-Dark) in different redshifts. Here we are trying to match the power spectrum of both TNG300-Dark and TNG300 (to account for the baryonic effects). The shadow region shows the $1\%$ deviation. The resolution of FastPM is 125 times lower than TNG300-2-Dark.

Here $k_l$ and $k_s$ are also free parameters.

[53] showed that the PGD model improves the halo profiles and the small-scale power spectrum. The PGD model can be treated as a single post-processing correction on the output snapshot, but it would be hard to do the correction on a lightcone output in this way, since the correction parameters are functions of redshift, and the redshift is not fixed in a lightcone output. Here we try to solve the problem by incorporating PGD correction into FastPM. We perform a PGD correction after each FastPM step, and then feed the corrected particle position into the next time step. Because the PGD is coupled into the simulation, both the static snapshots and the lightcone output are consistently corrected.

The parameters $\alpha$ and $k_l$ are functions of redshift, and we model their redshift dependence by

$$\log\left(\frac{\alpha}{\alpha_0}\right) = Aa^2 - Ba, \tag{3.14}$$

$$k_l = k_{l,0}a^\gamma, \tag{3.15}$$

where $\alpha_0$, $A$, $B$, $k_{l,0}$ and $\gamma$ are free parameters. $k_s = k_{s,0}$ is another free parameter and is fixed for all redshifts. These parameters are fitted by matching the matter power spectrum at all redshifts simultaneously. In figure 3.9 we show the matter power spectrum of original FastPM, and FastPM after the correction. Unlike [53] where they are comparing the same resolution simulation, here the mass resolution of FastPM is 125 times lower than the TNG300-2-Dark, yet we show that we can match the power spectrum quite well. The cross-correlation coefficient also improves on all redshifts, e.g., it improves approximately from 0.5 to 0.6 at the scale of $k = 10\,h^{-1}\,\mathrm{M}pc$.

## Light-cone simulation

To test how the PGD model improves the weak lensing map, We build a light-cone output from the FastPM simulation. The FastPM simulation is run in a $3200\,h^{-1}\,\mathrm{M}pc$ periodic box with $1536^3$ dark matter particles. The simulation starts at redshift 9, with time steps separated by constant spacing in the scale factor. The FastPM without PGD correction has 40 time steps, while after implementing the PGD correction we reduce the time steps to 20, since it has a comparable computation cost as a 40-step FastPM. The positions and velocities of the particles located between the steps are interpolated from the nearest steps and are saved as the particle positions intersect the observer's light-cone. An optional FoF halo finder can be run on the fly as the light-cone is generated, which uses padding to handle the continuity between light-cone slices. Given that the volume of the simulation box can be smaller than the light-cone, the simulation box is tiled (duplicated) as necessary to cover the required volume of the lightcone. The light-cone module of FastPM allows configurations on the position of the observer, the field of view angle ( determines the sky-area), the direction of sightlines, the replication (tiling) matrix, the list of culling octants, and the redshift range of interest.

In this work, we assume the observer sits at the origin of the simulation box, and integrates the sightlines up to $z = 2.2$. Note that the comoving distance to redshift 2.2 is around $3.8\,h^{-1}\,\mathrm{G}pc$, and the box is replicated along all directions to include the extra $600\,h^{-1}\,\mathrm{M}pc$.

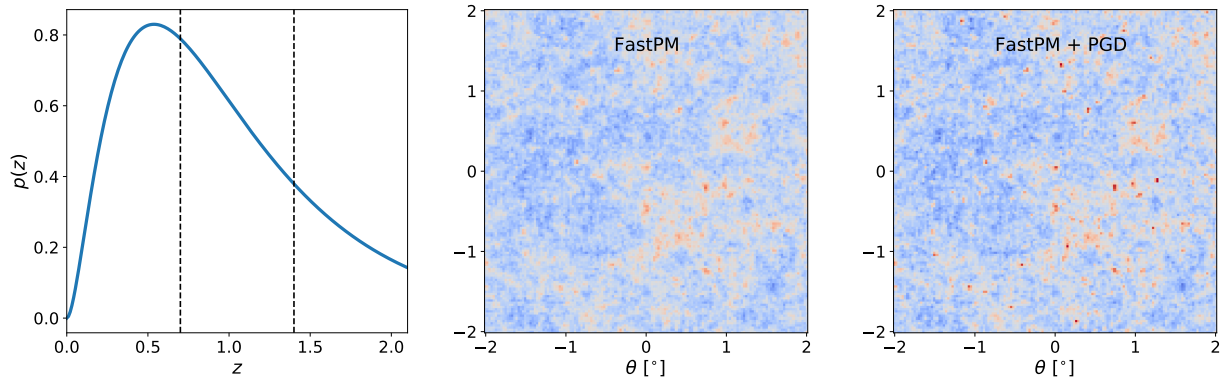



Figure 3.10: The redshift distribution of source galaxies of an LSST-like survey (left panel), and the zoomed-in convergence field of FastPM (middle panel) and FastPM with PGD correction (right panel).

## Weak lensing convergence

Under the Born approximation, we estimate the weak lensing convergence map produced by the source galaxies between redshift $z_{\rm min}$ and $z_{\rm max}$ as (see e.g. [134, 159])

$$\kappa(\theta) = \frac{3H_0^2\Omega_m}{2c^2}\int_{z_{\rm min}}^{z_{\rm max}} p(z_s)dz_s \int_0^{\chi_s(z_s)} d\chi \frac{(\chi_s(z_s)-\chi)\chi}{\chi_s(z_s)a}\delta(\chi,\theta) \tag{3.16}$$

where $\theta$ is a 2D angular vector, $\delta$ is the matter overdensity at radial comoving distance $\chi$ and angular position $\theta$, $\chi_s(z_s)$ is the comoving distance to redshift $z_s$, and $p(z_s)$ is the normalized redshift distribution of source galaxies between redshift $z_{\rm min}$ and $z_{\rm max}$. Weak lensing maps are generated in the post processing after the simulation has ended and the particle lightcone has been saved. In the post-processing step, the lightcone particles are read in, integrated along the line of sight using the weights based on lensing kernel, and then a pixelized map is generated with NGP (nearest grid point) window function using HEALPY[368], the python version of healpix [109]. While there is one I/O overhead due to lensing maps being generated in this manner, the post processing provides flexibility to generate multiple lensing maps for different lensing source configurations, and saving the lightcone also allows one to generate lightcones for different probes in general (the line of sight integration kernel can be different from the lensing kernel). Furthermore, for the case of cross correlations, it is also possible to generate the maps integrating over a narrow lens redshift range rather than over a complete redshift range from sources to observers, if necessary to reduce the I/O load.

In this work, we assume a source galaxy redshift distribution of an LSST-like survey (second panel of Figure 3.10). We divide the source into 3 tomographic bins: $z \in [0, 0.7]$, $[0.7, 1.4]$, and $[1.4, 2.1]$, and generate the convergence maps produced by these 3 source bins separately. In Figure 3.10 we show the all-sky convergence map as well as the zoomed-in maps of the last tomographic bin. We see that the PGD correction makes the peaks more evident. Thus we expect that the

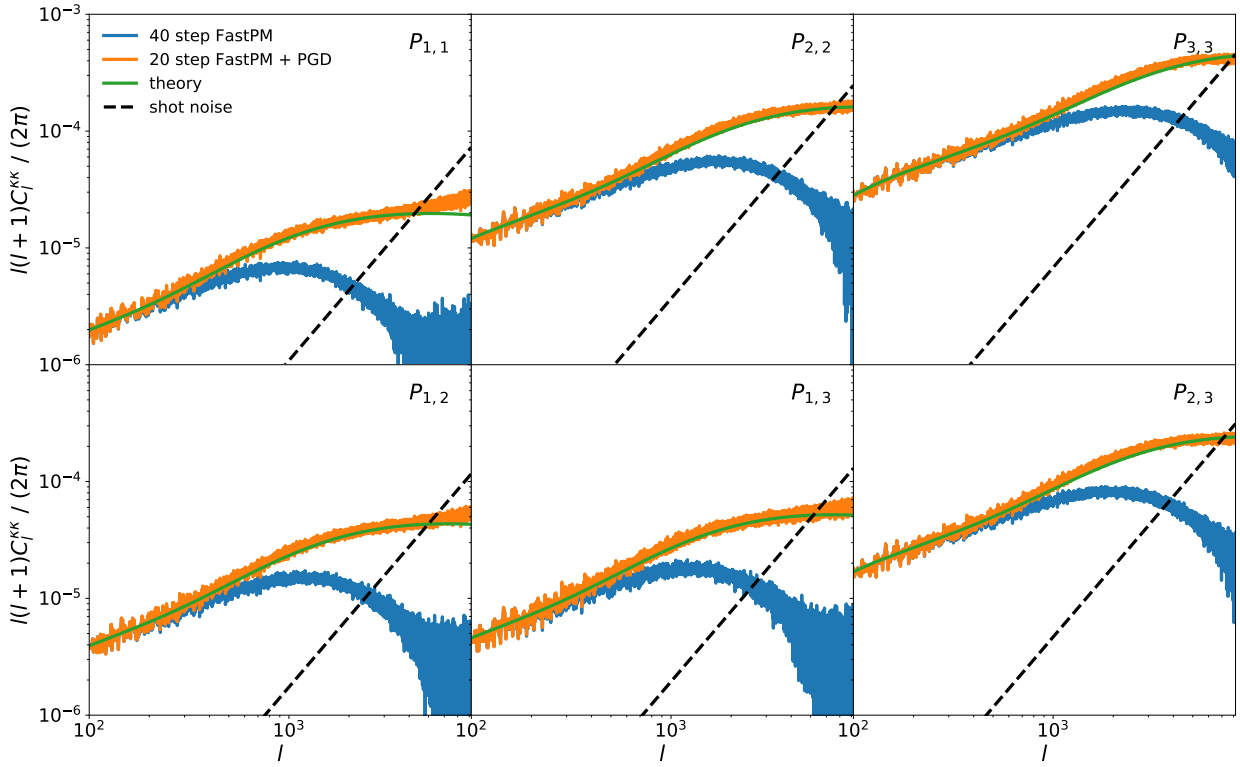



Figure 3.11: The weak lensing convergence auto power spectrum (upper panel) and cross power spectrum (bottom panel). The bin number 1, 2 and 3 correspond to tomographic bin $z \in [0, 0.7]$, $[0.7, 1.4]$, and $[1.4, 2.1]$, respectively. The theoretical weak lensing convergence power spectrum is calculated using Equation 3.17 with halofit nonlinear matter power spectrum. Black dashed lines show the shot noise power spectrum from the particles, which we subtract from the measured power spectrum.

correction to help with non-Gaussian statistics such as peak statistics. Below we will examine the auto power spectrum and cross-power spectrum of these convergence maps.

Under the Limber approximation, the angular power spectrum of the weak lensing convergence can be written as

$$C^{\kappa}(l) = \left(\frac{3H_0^2\Omega_m}{2c^2}\right)^2 \int_{z_{\min 1}}^{z_{\max 1}} p_1(z_{s,1})dz_{s,1} \int_{z_{\min 2}}^{z_{\max 2}} p_2(z_{s,2})dz_{s,2} \int_0^{\chi_s(\min(z_{s,1},z_{s,2}))} d\chi \left(\frac{\chi_s(z_{s,1})-\chi}{\chi_s(z_{s,1})a}\right)\left(\frac{\chi_s(z_{s,2})-\chi}{\chi_s(z_{s,2})a}\right) P_m(k=\frac{l+0.5}{\chi}, z(\chi)) \tag{3.17}$$

where $P_m(k, z)$ is the 3D matter power spectrum. We have assumed that $p_1(z)$ and $p_2(z)$ are normalized. $p_1(z)$ and $p_2(z)$ will be the same in the case of auto power spectrum, and different for the cross power spectrum. In Figure 3.11 we show the theoretical convergence power spectrum

calculated using halofit nonlinear matter power spectrum [310], as well as the power spectrum we measure using the simulated convergence map. After the PGD correction, the power spectrum matches the halofit predictions.

## 3.4 Conclusions

In this chapter, we improve the halo statistics and small-scale matter distribution in low-resolution fast quasi-N-body simulations. For halos, we introduce relaxed-FoF, a modification to the standard FoF algorithm so that the linking length is a function of the halo mass. For smaller halos, relaxed-FOF increases the linking length to enhance the identification of small halos, to improve agreement on the halo mass function, and to reduce the fraction of missed halos. We reject fake halos by reducing the linking length for the halos with large velocity dispersions. The rejection procedure removes fake halos found in the high-density regions, and therefore improves the halo bias. We find that using a high-resolution mesh for the 2LPT initial condition enhances the identification of small halos. We also calibrate the small-scale velocities of small halos by adding an irrotational velocity term. This extra term is written as the gradient of the velocity potential, which is learned from the gravitational potential. We verify the results on several halo statistics, including halo bias, halo mass function, halo auto power spectrum in real space and in redshift space, cross-correlation coefficient with the reference halo catalog, and halo-matter cross power spectrum. We find that our relaxed-FoF halo finder improves all of these. The ratio of missed halos and the halo catalog cross-correlation coefficient suggests that our halo catalog from FastPM is comparable to the halo catalog from a full N-body simulation of the same mass resolution, while our catalog has a better large-scale auto power spectrum in real space and redshift space, as well as better halo-matter cross power spectrum.

We also incorporate the potential gradient descent (PGD) method into FastPM simulation to improve the matter distribution at nonlinear scales. We couple the PGD correction into the FastPM time steps. We show that the fully coupled PGD correction improves the matter power spectrum measured from static snapshots at all redshifts, just as the previously studied static PGD method [53]. We build a light-cone simulation from a PGD-enabled FastPM simulation, by interpolating the particle location between the steps. We show that the PGD correction significantly improves the convergence tomographic power spectrum measured from the light-cone output.

There are several free parameters in relaxed-FOF and PGD. In principle, these free parameters depend on simulation resolutions, the number of steps, and potentially cosmological parameters. To achieve the best results, they need to be optimized for different situations. One could obtain the parameters by fitting them to a small volume high-resolution simulation with the same random seed. Here we do not try to study the parameter dependence besides the redshift dependence, which contains the dominant effect of amplitude dependence. While we expect the dependence on cosmological parameters other than amplitude to be small, this needs to be verified explicitly and is beyond the scope of this paper.

We plan to use FastPM for mock catalogs of both spectroscopic surveys such as DESI and photometric/weak lensing surveys such as LSST. The techniques we developed here will be useful

to improve the halo and matter statistics in those simulations, thus enabling one to simulate the whole survey at the required mass resolution. For example, for DESI one needs to resolve halos down to $10^{11}h^{-1}M_{\odot}$ and to cover the entire survey one needs volumes in excess of $(3\,h^{-1}\,\mathrm{G}pc)^3$, which can be achieved with $10^{12}$ particle FastPM simulations, similar to the one that has recently been run [223].

**Acknowledgements:** The majority of the computations were performed on NERSC computing facilities Edison and Cori, billed under the cosmosim and m3058 repository. National Energy Research Scientific Computing Center (NERSC) is a U.S. Department of Energy Office of Science User Facility operated under Contract No. DE-AC02-05CH11231. We thank Dylan Nelson and the IllustrisTNG team for kindly providing the linear power spectrum and random seed of the IllustrisTNG simulations. This material is based upon work supported by the National Science Foundation under Grant Numbers 1814370 and NSF 1839217, and by NASA under Grant Number 80NSSC18K1274.

## 3.5 Appendix A: Halo power spectrum of higher mass thresholds

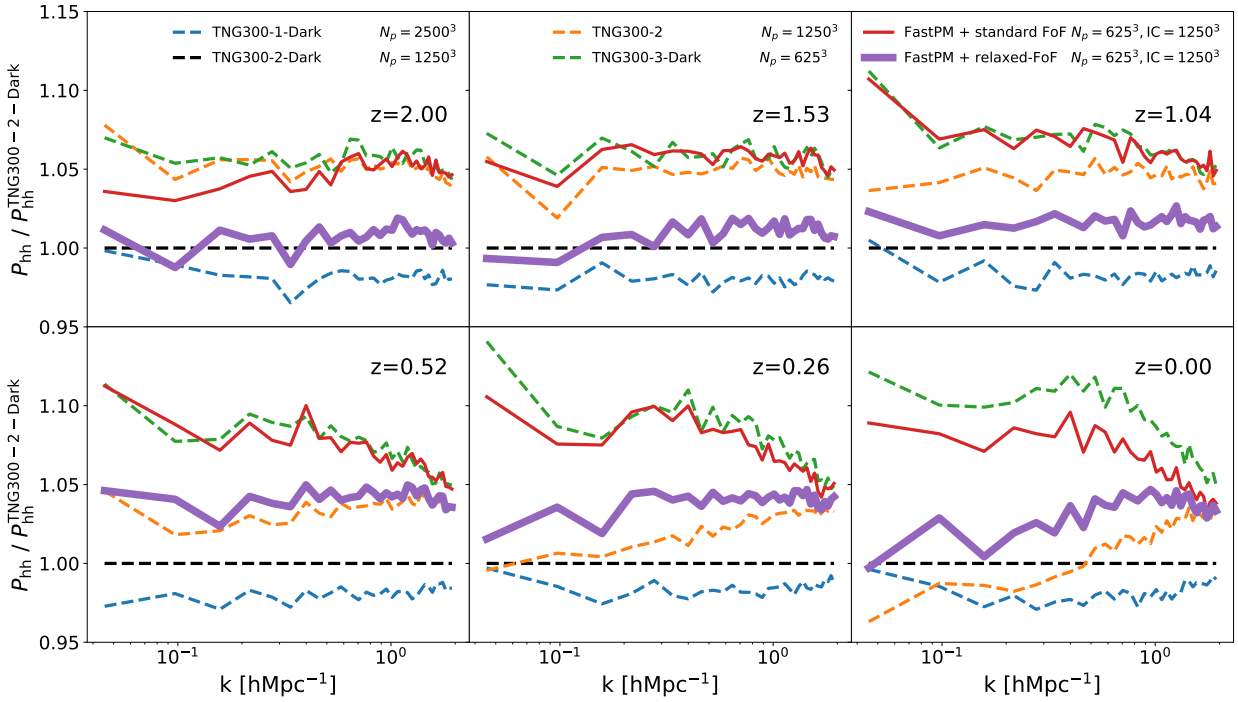



Figure 3.12: The ratio of halo auto power spectrum in real space from different simulations and various redshifts. The halos are selected using abundance matching, corresponding to $M \geq 2 \times 10^{11}M_{\odot} = 1.35 \times 10^{11}h^{-1}M_{\odot}$ halos (45 particles in FastPM).

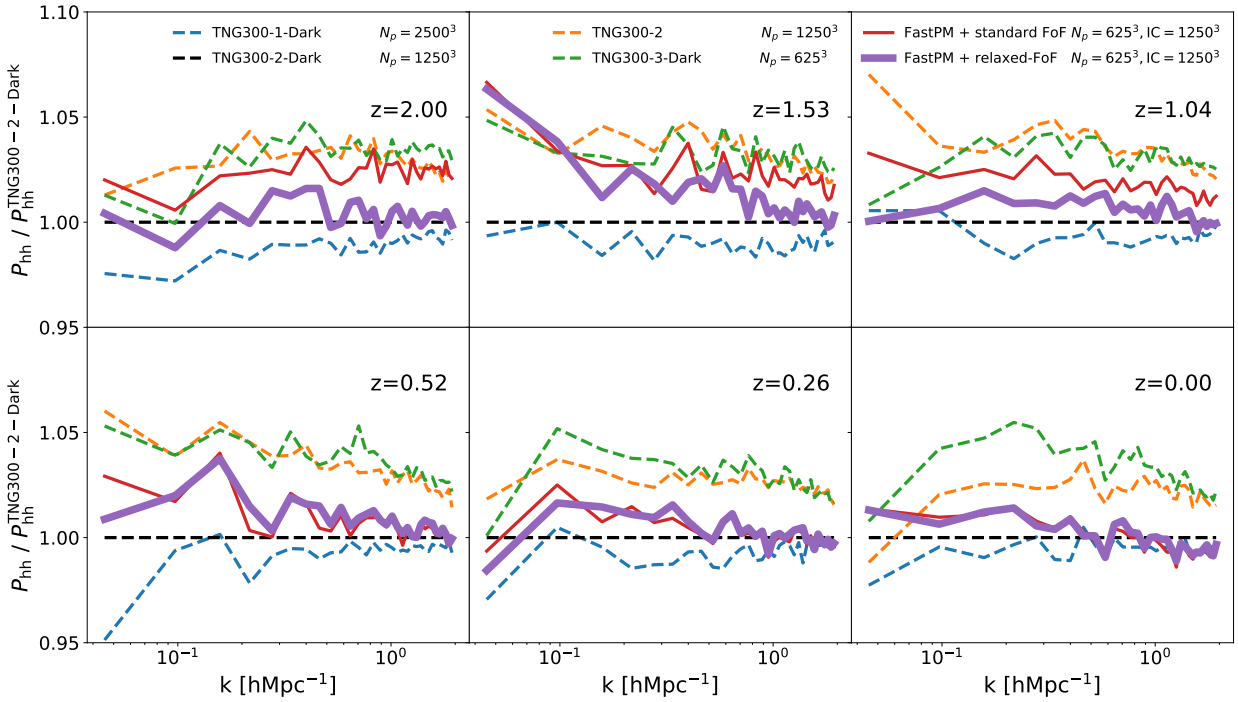



Figure 3.13: The same as Figure 3.12, but for halos $M \geq 10^{12} M_\odot = 6.8 \times 10^{11} h^{-1} M_\odot$. For this mass range (200 particle halos) we expect the predictions from FastPM to be accurate, and the difference between relaxed-FoF and standard FoF should be small.

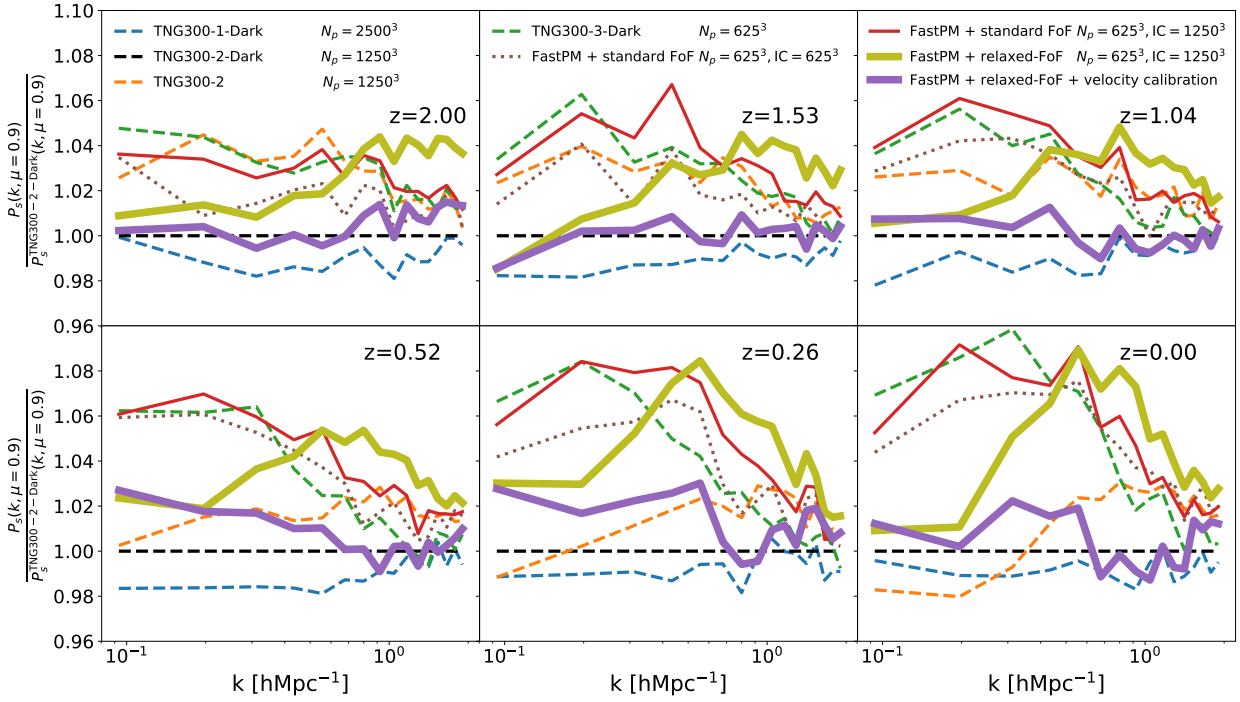



Figure 3.14: The ratio of halo auto power spectrum in redshift space from different simulations and various redshifts, for halos $M \geq 2 \times 10^{11} M_\odot = 1.35 \times 10^{11} h^{-1} M_\odot$. For RSD halo power spectrum with $\mu = 0$ see Figure 3.12.

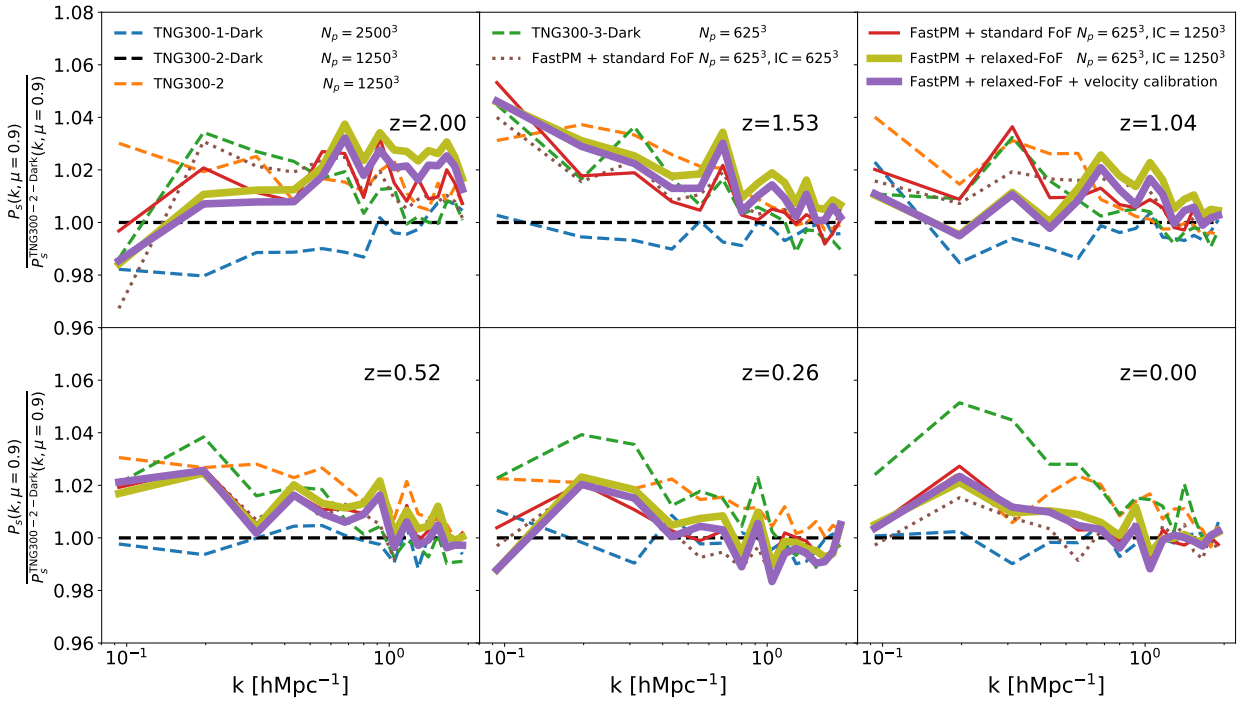



Figure 3.15: The same as Figure 3.14, but for halos $M \geq 10^{12} M_\odot = 6.8 \times 10^{11} h^{-1} M_\odot$. For RSD halo power spectrum with $\mu = 0$ see Figure 3.13. Again, we expect the predictions from FastPM to be accurate for this mass range (200 particle halos), and the difference between relaxed-FoF and standard FoF should be small.

# Chapter 4

# MADLens, a python package for fast and differentiable non-Gaussian lensing simulations

The contents of this chapter were originally published in [28],

> MADLens, a python package for fast and differentiable non-Gaussian lensing simulations
> *Böhm V., Feng Y., Lee M.E., Dai B.*(arXiv:2012.07266) A&C 36 (2021) 100490

We present MADLens, a python package for producing non-Gaussian lensing convergence maps at arbitrary source redshifts with unprecedented precision. MADLens is designed to achieve high accuracy while keeping computational costs as low as possible. A MADLens simulation with only $256^3$ particles produces convergence maps whose power agree with theoretical lensing power spectra up to $L{=}10000$ within the accuracy limits of HaloFit. This is made possible by a combination of a highly parallelizable particle-mesh algorithm, a sub-evolution scheme in the lensing projection, and a machine-learning-inspired sharpening step. Further, MADLens is fully differentiable with respect to the initial conditions of the underlying particle-mesh simulations and a number of cosmological parameters. These properties allow MADLens to be used as a forward model in Bayesian inference algorithms that require optimization or derivative-aided sampling. Another use case for MADLens is the production of large, high-resolution simulation sets as they are required for training novel deep-learning-based lensing analysis tools. We make the MADLens package publicly available under a Creative Commons License .

## 4.1 Introduction

Measurements of the weak cosmic shear signal will be among the major experimental drivers for advancing cosmology in the next decade. Next generation surveys such as the *Vera C. Rubin Observatory* [195], the *Roman Space Telescope* [298] and the *EUCLID* satellite [85] will provide an unprecedented amount of high resolution weak cosmic shear data, which creates a demand for

novel data analysis and modeling techniques. The weak cosmic shear signal is sensitive to the evolution of matter clustering over several orders of magnitude of scales, ranging from well within the linear to the highly non-linear regime. Tomographic lensing measurements are sensitive to the total matter content, $\Omega_{m0}$, the amplitude of clustering, $\sigma_8$, and the time evolution of clustering, which allows to constrain dark energy [137, 139, 297] and the sum of neutrino masses. Weak cosmic shear measurements can further be used to test general relativity [125, 280, 119].

Traditional lensing analyses mostly rely on two-point statistics or related observables [167, 129, 166, 6, 132, 64, 314, 131] to extract cosmological information. However, since the lensing convergence field is inherently and significantly non-Gaussian, two-point statistics do not exploit its full information content. In fact, a long list of studies have shown that non-Gaussian summary statistics, such as higher order correlation functions [248, 309, 145, 290, 98, 49] or peak statistics [144, 69, 211, 207, 254, 251, 38, 187, 192, 193, 155, 247, 186] can break parameter degeneracies that cumber power spectrum analyses and lead to significantly tighter constraints. While inference from these summaries offers improvements over power spectra analyses, their choice is somewhat ad hoc and the question of how to best extract cosmological information from non-Gaussian lensing scales is still an active field of research.

A number of works have recently suggested machine-learning tools for identifying informative summaries [112, 270] and even successfully applied them to real data [92, 149]. Machine learning methods generally require a large amount of training data. If applied correctly, these methods extract features in the lensing map which are highly informative about cosmological parameters. A recent study finds that these models are mostly sensitive to extreme values in the lensing field [210]. This underlines the importance of training data that accurately mimics real cosmic shear data and its dependence on cosmological parameters down to very small scales.

Another approach, and in principle the optimal one, is to build a differentiable non-linear data model that starts from the Gaussian initial conditions and forward models them accurately to the measured lensing signal. This forward model is used to model the posterior of the parameters of interest (these can be cosmological parameters, the modes of the initial field or the bandpowers of the initial power spectrum). Analyzing this posterior generally relies on powerful sampling or optimization schemes, which in turn require many model evaluations and often the derivatives of the model with respect to the parameters of interest. Forward-model based inference schemes have been developed for a range of of observables in cosmology [288, 287, 146, 147] including weak cosmic shear [29, 258].

All of these new avenues for lensing analyses create the need for fast and differentiable simulations of the lensing field that at the same time accurately capture the nonlinear features of that field.

Realistic lensing simulations are challenging because a range of scales in the three dimensional matter distribution contribute to a single angular scale in the lensing field making even intermediate lensing scales sensitive to the non-linearity of structure formation on small scales [143]. Accurate lensing simulations rely on lightcones constructed from high resolution N-body simulations. These N-body simulations must accurately resolve small scales, but must at the same time be large enough to produce lensing maps with an extent of several degrees.

A number of recent works have studied the applicability of deep generative models, in particular

generative adversarial networks (GANs) for producing accurate lensing convergence maps at low computational costs [227, 249]. These models do not aim at simulating the underlying physics, but are trained to mimic the training data to a degree where their output becomes indistinguishable from the training data for a neural network. While these early studies look promising, future research will have to show that these models indeed learn the correct data distribution or that using their outputs for inference leads to unbiased parameter posteriors. Another, more safeguard approach is to use machine-learning inspired techniques to boost the accuracy of low resolution N-body simulations [58, 54] and to construct lightcones from those. This is the avenue we have chosen in this work to create high resolution lensing simulations from approximate N-body solvers.

In this publication we describe a new, weak gravitational lensing package, MADLens. MADLens is a python package that allows one to compute fully nonlinear lensing convergence maps at different source redshifts and low computational cost while accurately modeling the non-Gaussianity of the field down to scales of several tens of arcseconds. MADLens is built on top of a particle-mesh solver that evolves an initial linear density field into non-linear late time density fields. It provides derivatives with respect to the initial conditions of the particle-mesh simulation and a number of cosmological parameters through automated differentiation. MADLens fills the gap in the accuracy-speed space between computationally expensive, high accuracy lensing simulations and fast approximate simulations. In particular, MADLens correctly captures scales down to $L{=}10000$ at a field of view (FOV) of $6.2^\circ$ with percent level precision at a runtime of 30 seconds with 32 processes. MADLens can be run at different levels of resolution and approximations, allowing the user to choose the speed to accuracy trade-off that is optimal for their application.

We begin this paper with a brief introduction of the cosmological-scale weak lensing formalism and our notation in Section 4.2. This is followed by a in depth discussion of the package design in Section 4.3. We demonstrate the packages abilities in a number of tests in Section 4.4 and conclude with a summary and outlook in Section 4.5. 4.6 provides details on the novel feature of differentiability with respect to cosmological parameters.

## 4.2 Weak Gravitational Lensing, Notation and Conventions

Weak gravitational lensing observations provide insight into the projected matter density distribution between an observer and a source through correlated image distortions. Here, we provide a brief overview and define our usage of the lensing kernel, lensing convergence, and power spectrum of cosmic shear used throughout MADLens (for a detailed discussion of weak gravitational lensing and especially cosmic shear, see [16, 159, 17]).

The image of a source galaxy at a comoving distance $\chi_s$ is distorted along the line of sight by some lensing potential $\Psi$. The potential of an extended lens under the Born approximation representing all density fluctuations along a line of sight at some angular position $\vec{\theta}$ can be found by integrating individual Weyl potentials $\Phi$ up to the comoving distance of the galaxy,

$$\Psi(\vec{\theta}) = \frac{2}{c^2}\int_0^{\chi_s} d\chi \frac{\chi_s - \chi}{\chi_s \chi}\Phi(\chi_{\vec{\theta}}, \chi), \tag{4.1}$$

where $\chi_{\vec{\theta}}$ is the angular perpendicular component of the potential and $\chi$ is the parallel component. The lensing deflection and convergence are defined as,

$$\begin{aligned} \vec{\alpha} &= \vec{\nabla}\Psi, \\ 2\kappa &= \nabla^2\Psi, \end{aligned} \tag{4.2}$$

where the derivatives are taken with respect to $\vec{\theta}$. Using the Poisson equation and neglecting the derivative along the line-of-sight direction, the lensing convergence can be rewritten as,

$$\kappa(\theta) = \frac{1}{2}\nabla^2\Psi(\theta) = \frac{3H_0^2\Omega_{m0}}{2c^2}\int_0^{\chi_{\rm lim}} \frac{d\chi}{a(\chi)} q(\chi)\delta(\chi\theta,\chi) \tag{4.3}$$

$$q(\chi) = \int_\chi^{\chi_{\rm lim}} d\chi' n(\chi') \frac{\chi(\chi'-\chi)}{\chi'}, \tag{4.4}$$

where $\delta$ represents the density contrast from the mean density, $H_0$ is the Hubble parameter and $q$ is the lensing kernel describing the projection of sources selected by the redshift selection function $n(\chi')$. MADLens evaluates the integral in Eq. 4.3 numerically. The most commonly used summary statistic in lensing analyses that can also be computed analytically is the power spectrum. To compute the convergence power spectrum we use Limber's approximation and the flat sky approximation which are both valid on intermediate and small scales,

$$C_L^{\kappa\kappa} = \left[\frac{3H_0^2\Omega_{m0}}{2c^2}\right]^2 \int_0^{\chi_s} d\chi \left(\frac{\chi_s-\chi}{\chi_s}\right)^2 P_m\left(k=\frac{l+0.5}{\chi}, z(\chi)\right). \tag{4.5}$$

Since the lensing power spectrum is very sensitive to non-linear corrections to the matter power spectrum, we use HaloFit [310] to model $P_m(k)$ throughout this paper.

## 4.3 MADLens package design

The MADLens package is based on FastPM [89], a highly scalable particle-mesh solver, that evolves particle positions through a kick and drift scheme enforcing correct linear displacement in each step. FastPM has been implemented in C and Python and two versions of FastPM support automatic differentiation, including the MPI based version used in this work, and the recently published FlowPM [222] package, which is based on TensorFlow. A FastPM particle-mesh simulation requires the choice of a particle-mesh resolution, equivalent to the number of particles in the simulation, the force resolution, the resolution of the grid onto which the particles are painted to compute the forces, the box size of the simulation and the number of steps in which the particle positions are evolved. The initial conditions and particle evolution depend on the cosmological parameters $\sigma_8$ and $\Omega_{m0}$.

MADLens runs a FastPM simulation. As the simulation evolves, MADLens projects the particles in the simulation weighted by the lensing kernel at each simulation step to 2D meshes at the desired source redshifts. The field of view, i.e. the size of the convergence map, and its

| Parameter | Description | Typical Value(s) |
|---|---|---|
| BoxSize | side length of the simulation box | 128-1024 Mpc/h |
| Nmesh | resolution of the particle-mesh simulation | $64^3 - 512^3$ |
| B | force resolution factor | 2 |
| Nsteps | number of steps in the FastPM simulation | $11 - 40$ |
| N_maps | number of output maps | $\geq 1$ |
| Nmesh2D | resolution of the convergence map | $256^2 - 2048^2$ |
| BoxSize2D | size of the convergence map in degrees | $2.5° - 22°$ |
| zs_source | list of source redshifts | $0.3 - 2.0$ |
| Omega_m | total matter density | 0.32 |
| sigma_8 | amplitude of matter fluctuations | 0.82 |
| PGD | whether to use PGD enhancement or not | True/False |
| interpolation | whether to use the sub-evolution scheme | True/False |

Table 4.1: List of MADLens simulation parameters that can be set by the user and their typical values.

resolution can be set by the user given the simulation box covers the entire field of view at the most distant source redshift. Each FastPM step evolves particles in redshift steps, $\Delta z_i = z_{i+1} - z_i$. MADLens constructs the lightcone by translating these redshifts into distances $\Delta\chi_i = \chi_{i+1} - \chi_i$ and projecting particles at the correct evolution step corresponding to that distance onto the convergence map. If the distance between two simulation steps is larger than the extent of the box, the box is replicated at the same redshift as often as is needed to fill the entire extent. In order to avoid spurious correlations, the simulations box is rotated before being repeated. In these techniques, MADLens is constructed similarly to other lightcone packages. We provide an overview of all MADLens parameters that can be set by the user in table 4.1.

To reach extraordinary accuracy at low computational costs, MADLens employs two special techniques:

- Particle Gradient Descent (PGD) [53] is an additional particle evolution step that corrects for the difference between particle distributions in a low resolution simulation and a high resolution simulation. The correction is applied after each simulation step. PGD introduces 4 additional nuisance parameters, which are fitted on training simulations. PGD allows simulations to run at lower resolution while still obtaining results that are comparable and highly correlated with a high resolution simulation. For details we refer the reader to chapter 2.

- A sub-evolution step allows for a massive reduction of the number of simulation steps. When using sub-evolution, particles are evolved according to the redshift of their position within the simulation box before projection, rather than by the redshift of the FastPM step. We provide more details on this scheme in 4.7.

Finally, MADLens is made differentiable through numerically accurate tape-based automatic differentiation. Specifically it uses the Virtual Machine Automated Differentiation package (VMAD). VMAD builds a computational operation graph that is traversed for the model evaluation, during which all operations are recorded on a sequential tape. Gradient graphs generated from the tape are used to compute Jacobian vector products ($J_{ij}v_j$) and vector Jacobian products ($v_iJ_{ij}$, commonly referred to as back-propagation). MADLens is made available in two variants. In its main version it is built to provide differentiability with respect to the initial, Gaussian modes of the simulation. A second package version, that is included in this release, adds differentiability with respect to the cosmological parameters, $\Omega_{m0}$ and $\sigma_8$. Differentiability with respect to the PGD parameters ($k_l$, $k_s$, $\alpha_0$, $\mu$) will be included in a future release.

## 4.4 Results

We analyze the performance of MADLens, the PGD enhancement, the sub-evolution scheme and computation times as well as the accuracy of the gradient computation. Testing the accuracy of MADLens output is challenging, because of the lack of a ground truth. We will use theoretical power spectra based on Halofit matter power spectra and high resolution runs for comparison. These can serve as a reasonable baseline, but as should become evident from our analysis, should not be mistaken for the ground truth.

In Figure 4.1 we show an example of a convergence map produced with MADLens. The non-Gaussian structure is clearly visible by eye and also evident in the corresponding histogram in Figure 4.2. The map resolves sub arcmin scales and extends over more than $6^\circ$ on the sky showing that MADLens overcomes one of the key challenges in lensing simulations: the accurate modeling of both large and small scales.

We compare the histogram of pixel values in Figure 4.2 with a log-normal distribution. Log-normal distributions have been used in the past to model lensing PDFs, but are not strictly theoretically motivated. The log-normal fit captures the rough shape of the distribution, but underestimates the probability of high values and overestimates the probability of low values.

### Accuracy

In Figure 4.3 we compare the power spectra measured from MADLens outputs at different source redshifts with the analytical model of Eq. 4.5 based on a HaloFit matter power spectrum. For this comparison we average the power spectra of five simulations to reduce the variance. Overall we find that the MADLens power spectra trace the theoretical predictions well within $10\%$ up to scales of a few thousand. At very small scales shot noise starts to contribute significantly to the power. To put the importance of this shot noise into perspective we further plot the experimental noise level expected in a typical future experiment, such as LSST, and find that the shot noise is subdominant to the expected noise levels in real data.

Figure 4.4 delves further into the comparison with theoretical power spectrum and quantification of the shot noise. We translate the HaloFit accuracy ($5\%$ for $k \leq 1\ \mathrm{h\,Mpc^{-1}}$ at $0 \leq z \leq 10$ and

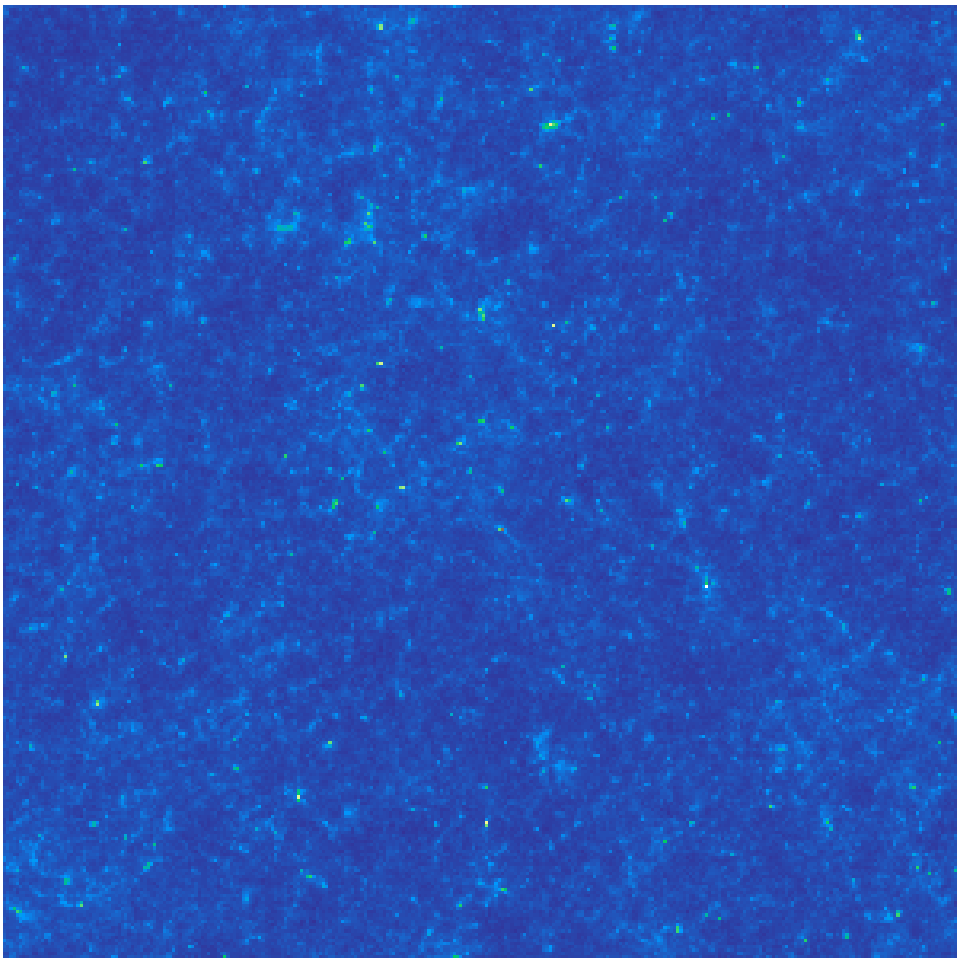


Figure 4.1: A MADLens convergence map at $z$=1.0, based on a 3D simulation of side length $256\ \mathrm{Mpc/h}$ and $256^3$ particles. The 2D lensing map has an angular extent of $6.2°$. It was down-sampled to a map of pixel size $43\ \mathrm{arcsec}$ and Nyquist frequency $L$=15000. The Non-Gaussianity is clearly visible by eye.

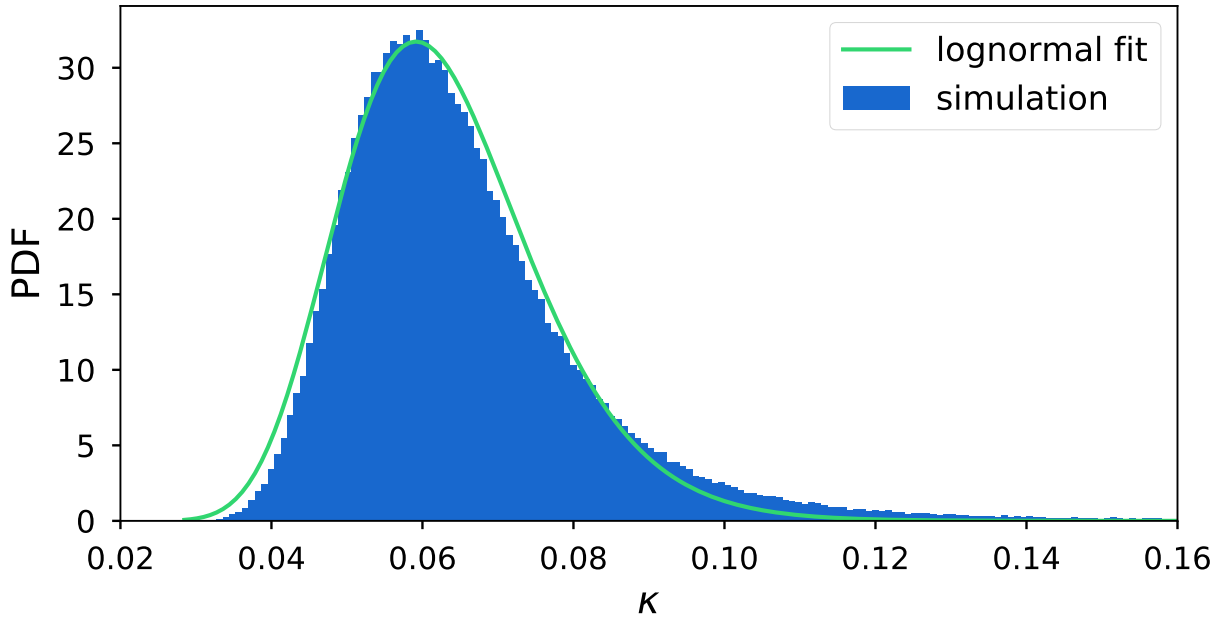



Figure 4.2: PDF of convergence values in the map in Figure 4.1. The distribution is clearly non-Gaussian with a pronounced skewness. A lognormal $\chi^2$ fit to the distribution (green) approximates its shape to some extent but does not accurately capture the tails.

$10\%$ for $1 \leq k \leq 10\ \mathrm{h\,Mpc^{-1}}$ at $0 \leq z \leq 3$) into accuracy in the lensing power spectrum and show these intervals as gray bands. The MADLens power spectrum lies well within these bands up to wavenumbers of a few thousand, where it becomes dominated by shot noise. We estimate the shot noise level by running a number of MADlens simulations with random particle positions (dark gray line) and subtract the result from the MADLens power spectrum (dark blue). The result lies within the HaloFit accuracy up to $L$=10000.

Figure 4.5 shows that the PGD enhancement allows to reach these high accuracies at much lower computational cost than conventional lensing simulations. We compare the output of MADLens simulations at a resolution of 1 particle per Mpc/h cubed to a conventional simulation (MADLens without PGD) with an eight times higher resolution. The higher resolution simulation not only requires about eight times more memory, but also takes more than twice as long. The high resolution run is of comparable resolution to other state-of-the-art lensing simulations which have been used for cosmological parameter inference studies [191], but the lower resolution MADLens simulation traces the theoretical convergence power spectrum up to much higher wavenumbers.

In Figure 4.6 we show cross correlations defined by

$$r(L) = \frac{C_L^{XY}}{\sqrt{C_L^{XX} C_L^{YY}}}, \tag{4.6}$$

where $X$ is the high resolution run without PGD enhancement, and $Y$ are lower resolution MADLens

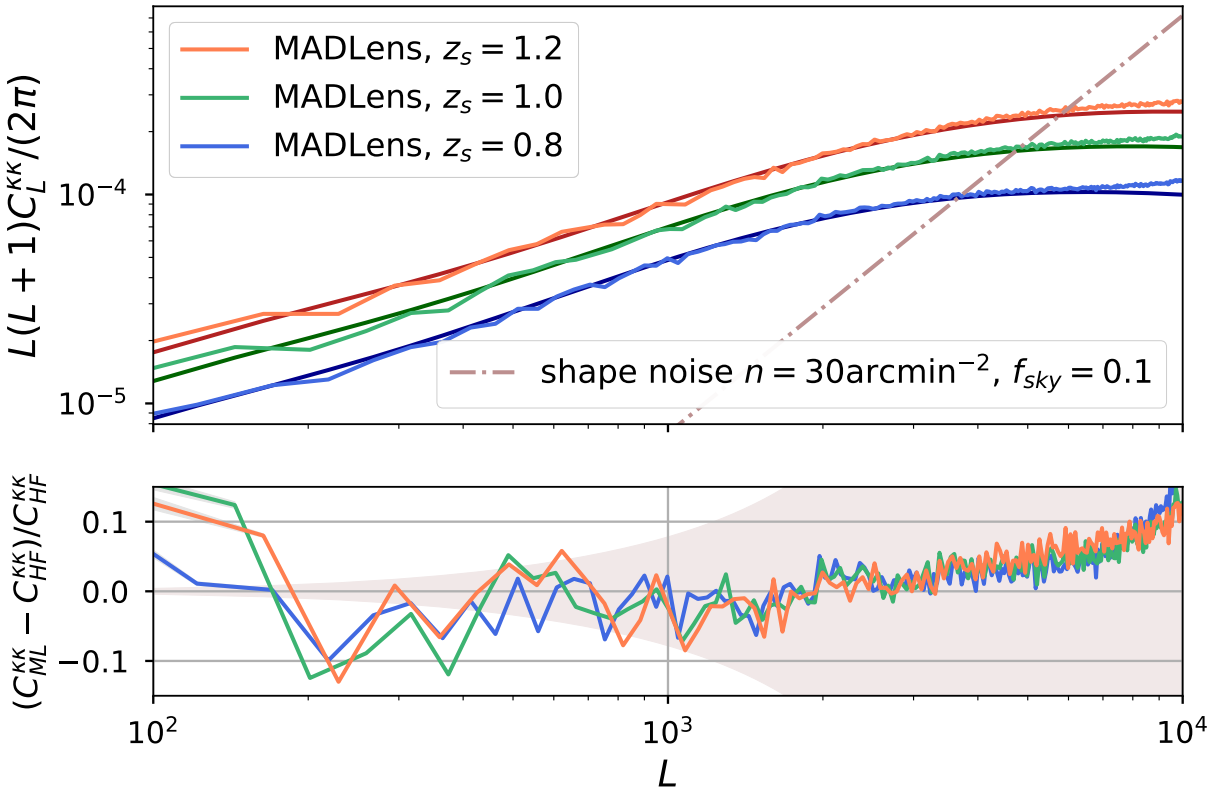



Figure 4.3: MADLens outputs (box length 512 Mpc/h, $512^3$ particles, with sub-evolution and PGD enhancement) for different source redshifts but same initial conditions (no shot noise subtraction). The measured power spectra agree with theoretical predictions up to very high wavenumbers independent of the source redshift. Lower source redshifts show slightly higher shot noise due to the lower number of particles that contribute to the projection. For comparison we plot the experimental noise for a typical galaxy density expected for future lensing experiments (pink dashed-dotted line and shaded area), showing that areas with significant shot noise contribution lie well within the experimental noise dominated regime.

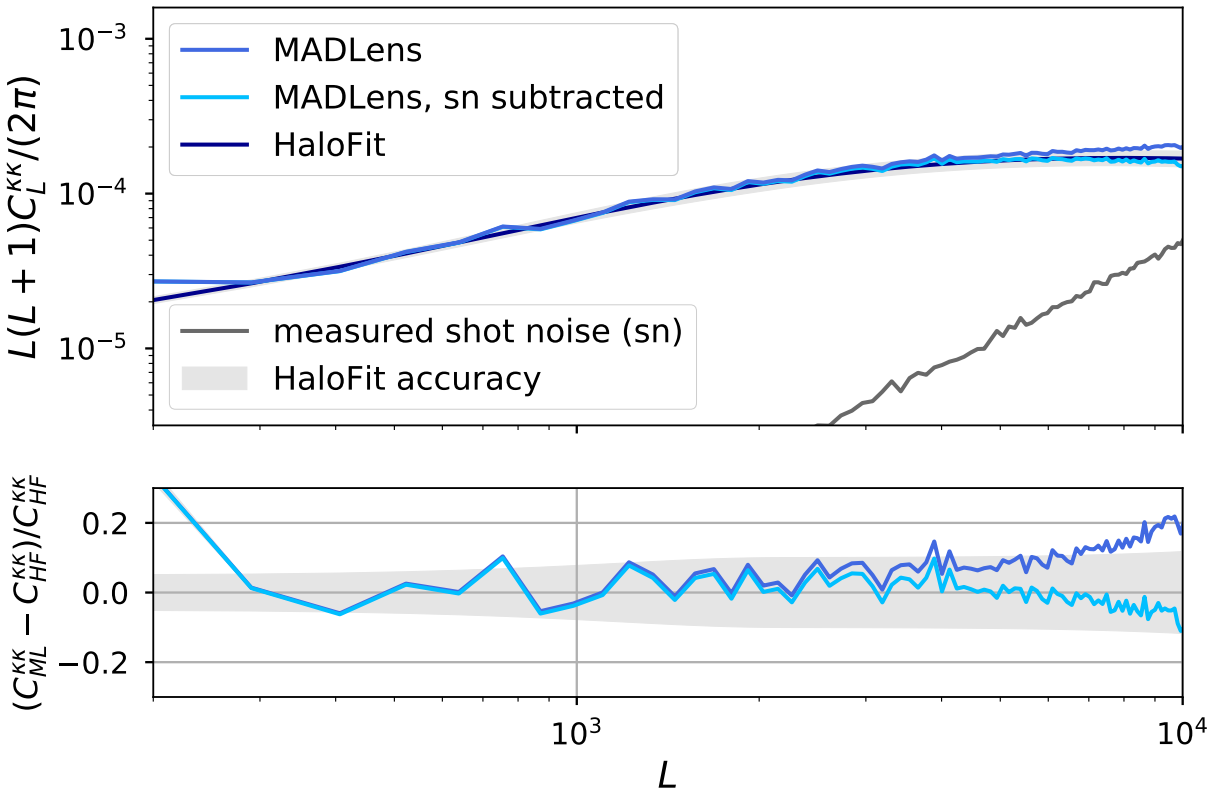



Figure 4.4: Shot noise and comparison to a theoretical convergence power spectrum based on a HaloFit matter power spectrum: The MADLens simulation traces the HaloFit power spectrum within the accuracy of HaloFit (gray band). At high wavenumbers ($L$>4000) the shot noise (dark gray line) that is due to the small number of particles in the simulation ($256^3$) starts to contribute significantly to the signal. After subtraction of the shot noise, the convergence power lies well within the uncertainty band up to $L$=10000.

outputs that have either been produced with or without PGD enhancement. As expected, the PGD enhanced lower resolution map shows higher correlation (dark blue) with the high resolution run than the one without enhancement (light blue) on intermediate scales. The results of this cross correlation analysis must be taken with a grain of salt: since the high resolution run is suffering from a significant lack of power on small scales, this could also be an indication of inaccurate particle positions. A lower cross correlation on small scales simply states that the simulations differ significantly, but does not show which one is more correct.

The accuracy of MADLens is further boosted by a sub-evolution projection scheme, where particles are moved to the position corresponding to their actual distance to the observer before being projected on the lensing map. We illustrate the efficacy of this scheme in Figure 4.7. An 11-step simulation naturally overestimates the total lensing power (light blue). The sub-evolution

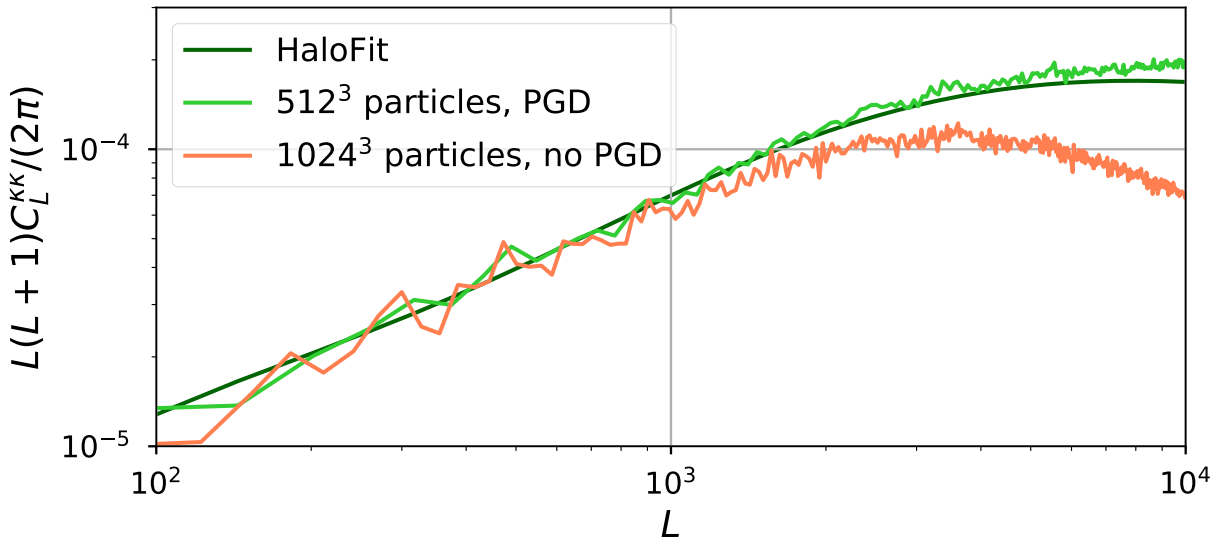



Figure 4.5: The PGD enhancement recovers lensing power on small scales in a low resolution simulation: We show a comparison between the convergence power spectra of a MADLens simulation with $512^3$ particles in a box of side length 512 Mpc/h that was run with PGD enhancement (light green) and an otherwise identical simulation with 8 times more particles that was run without PGD (orange). The latter corresponds to a setting used in some state-of-the-art simulations. The PGD enhanced simulation tracks the theory power spectrum (dark green) well beyond $L$=1000, while the standard simulation suffers from a significant loss in power on small scales.

scheme is able to correct for this overestimation up to scales $L$>2000 (dark blue). In a 40 step simulation the discrepancy between actual particle positions and their true evolution stage is smaller (light red), however, even the accuracy of a 40 step simulation can be enhanced by the sub-evolution scheme (dark red).

The MADLens derivatives have been thoroughly tested and verified with VMAD built-in test functions. Here we show that the derivatives are accurate by means of a single example: we build a finite difference test by slightly changing a single pixel value in the initial field. We then take the difference of the output maps generated from runs with slightly different values in this initial pixel and compare it with the output of the Jacobian-vector-product (Jvp), where the vector encapsulates the change in the initial field. If the Jvp vector product is correct, the result should agree with the difference of the output maps. That this is indeed the case is shown in Figure 4.8. The first two panels show the Jvp and the finite difference result, respectively. They are indistinguishable by eye. The next panel shows the difference between the first two panels, revealing insignificant numerical inaccuracies, five magnitudes smaller than the signal. In the last panel we compare the outputs in terms of their histograms, finding again an excellent match.

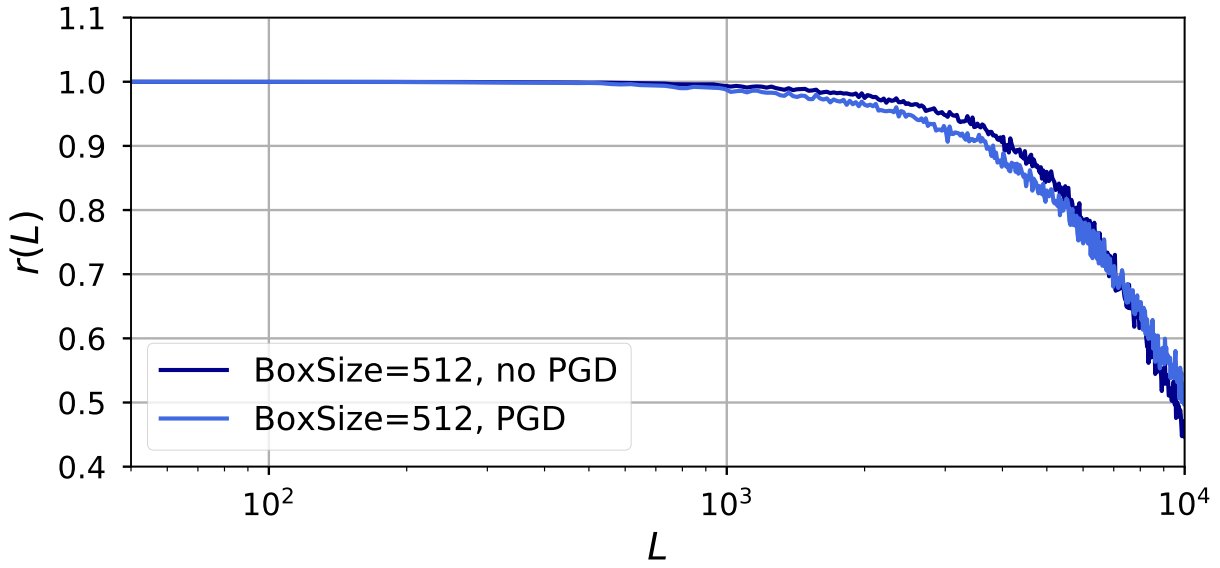



Figure 4.6: Cross correlation between a low resolution MADLens run with and without PGD enhancement, and a simulation with 8 times more particles and no PGD enhancement (same as Fig 4.5). The PGD enhancement increases the cross correlation on intermediate scales demonstrating that PGD moves particles in a physically sensible way. Note that the high resolution simulation is not the *truth* as it lacks a significant amount of power on small scales.

## Run times

We conducted timing tests for MADLens on Intel® Xeon® Processors E5-2698 v3 (NERSC Cori Haswell nodes), and show results in Figures 4.9-4.12. A single MADLens simulation that achieves accuracies as shown in the last section takes of the order of 10-60 seconds with 32 processes. The scaling of the run-time with source redshift is roughly linear and reducing the number of particles by a factor of 8 reduces the run-time to about one third (Figure 4.9).

The computation time can be further reduced by parallelizing with up to $128$ processes, after which the communication overhead starts to dominate the time budget (Figure 4.10).

Reducing the number of FastPM steps leads to significant savings in run time as we demonstrate in Figure 4.11. A conventional lightcone code requires about 40 steps in order to reach percent accuracies in the lensing power spectrum up $L{\approx}1000$. With PGD enhancement and sub-evolution scheme, MADLens reaches percent accuracies up to $L{=}2000$ with only 11 FastPM steps: a factor of 3 in time-savings.

The use of back-propagation to calculate the derivatives results in run times that are similar to the forward model. This is shown in Figure 4.12, where we find run times of 1.1-1.6 times the run time of the forward model for either Jvp and vJp.

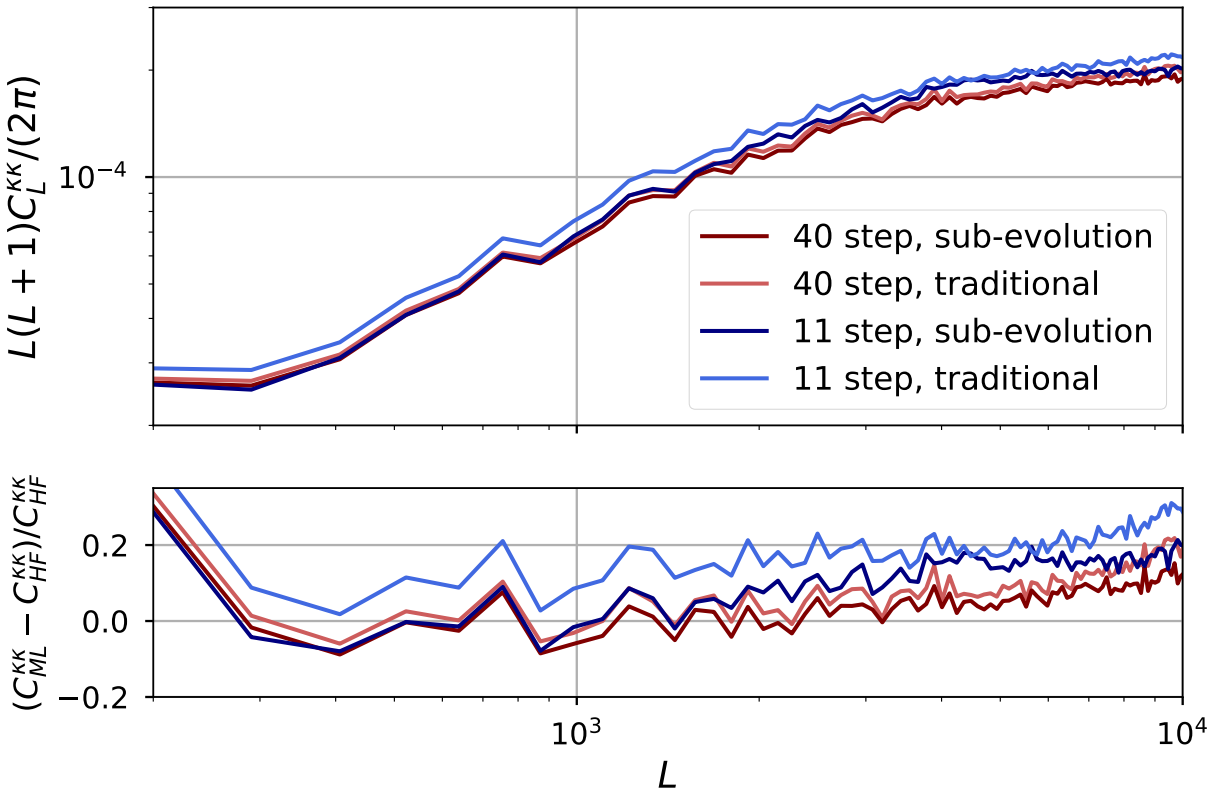



Figure 4.7: The sub-evolution scheme corrects for a systematic overestimation in power that occurs when snapshots are projected at the end of every FastPM step instead of at the exact position/redshift that corresponds to their distance to the observer (light blue). While the discrepancy can be reduced by using more simulation steps and hence higher computational cost (light red), a sub-evolution step before each projection significantly improves the output at a fixed number of simulation steps (dark blue and red).

## 4.5 Summary & Outlook

We have presented MADLens, a fully differentiable python package for producing non-Gaussian convergence maps of weak gravitational lensing on cosmological scales. MADLens reaches unprecedented accuracy even when compared to many non-differentiable lensing simulations, and operates at run times of a factor of 2-20 below conventional N-body simulation based lightcone packages. These advancements are made possible by several features, including a machine learning inspired post processing step, that allows the N-body simulation to run at a lower resolution and with less steps without paying a significant penalty in accuracy. The post-processing step is a physically motivated model which approximates the particle positions in a high resolution simulation. It introduces 4 additional nuisance parameters which are calibrated against training simulations.

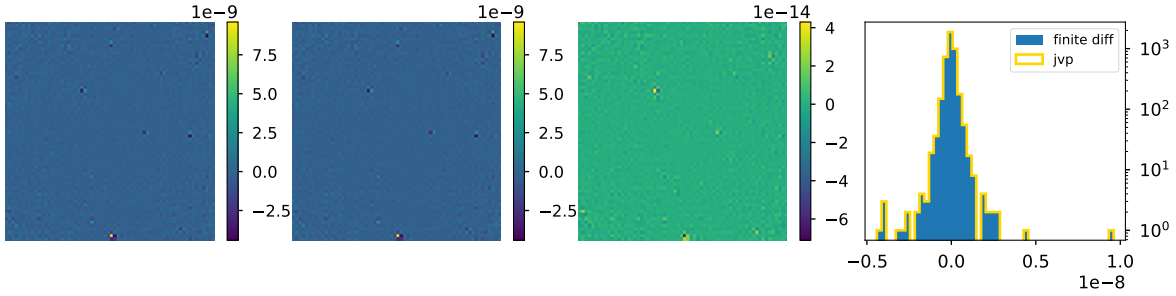



Figure 4.8: The MADLens automatic derivative agrees excellently with the result of finite differencing. To show this here, we measure the response of the convergence map by slightly changing the initial field in a single pixel. The corresponding MADLens Jacobian-vector-product in the first plot and the finite difference result in the second plot agree to the order of $10^{-5}$, as can be seen in the difference map (third plot) and by comparing their histograms (fourth plot).

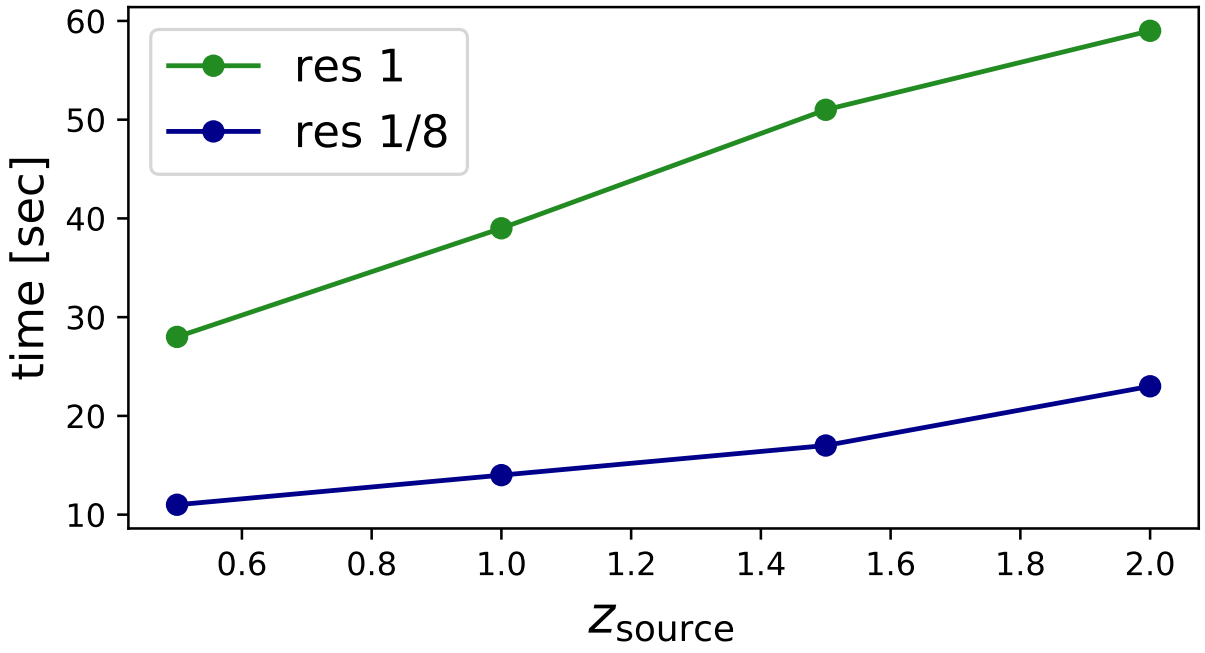



Figure 4.9: Computation time for a single simulation run with 11 steps and a box size of 256 Mpc/h for different source redshifts. We compare a run with $256^3$ particles ($\text{res} = 1\,[\text{h/Mpc}]^3$) particles to a run with $128^3$ particles ($\text{res} = 0.125\,[\text{h/Mpc}]^3$). All simulations were run on a single node with 32 processes in this test.

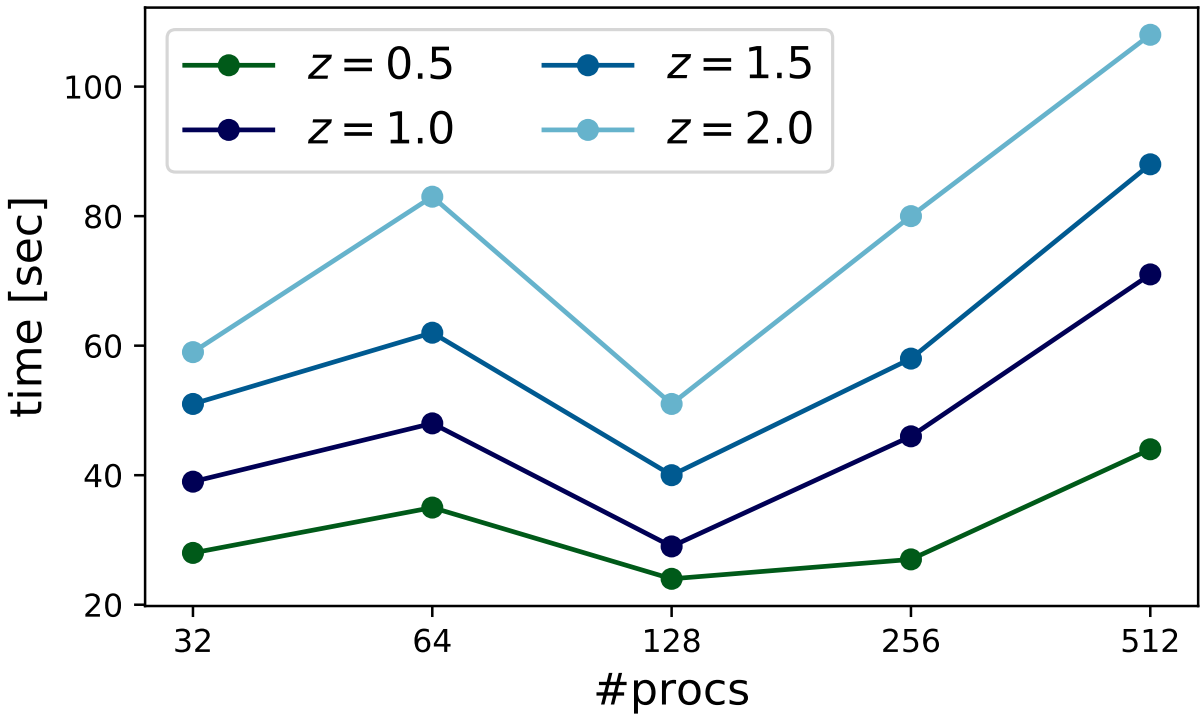



Figure 4.10: Scaling of the computation time with the number of processes for a $256^3$ particle/ 256 Mpc/h box simulation. Different lines represent different source redshifts.

Taking the derivative through a MADLens simulation with respect to the initial modes of the N-body simulation and the two key cosmological parameters $\sigma_8$ and $\Omega_{m0}$ is made possible through back-propagation. This means that evaluating the derivatives has comparable computational cost as the forward simulation. With these features MADLens constitutes a milestone towards the development of fully differentiable inference pipelines for weak cosmic shear. In the future MADLens will be integrated into the tensorflow-based FlowPM framework. Package updates will also feature differentiability with respect to nuisance parameters, such as the PGD parameters.

In the interest of scientific advancement and reproducibility, we make the MADLens package publicly available on github [1].

## Acknowledgements

This research used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility operated under Contract

[1] https://github.com/VMBoehm/MADLens

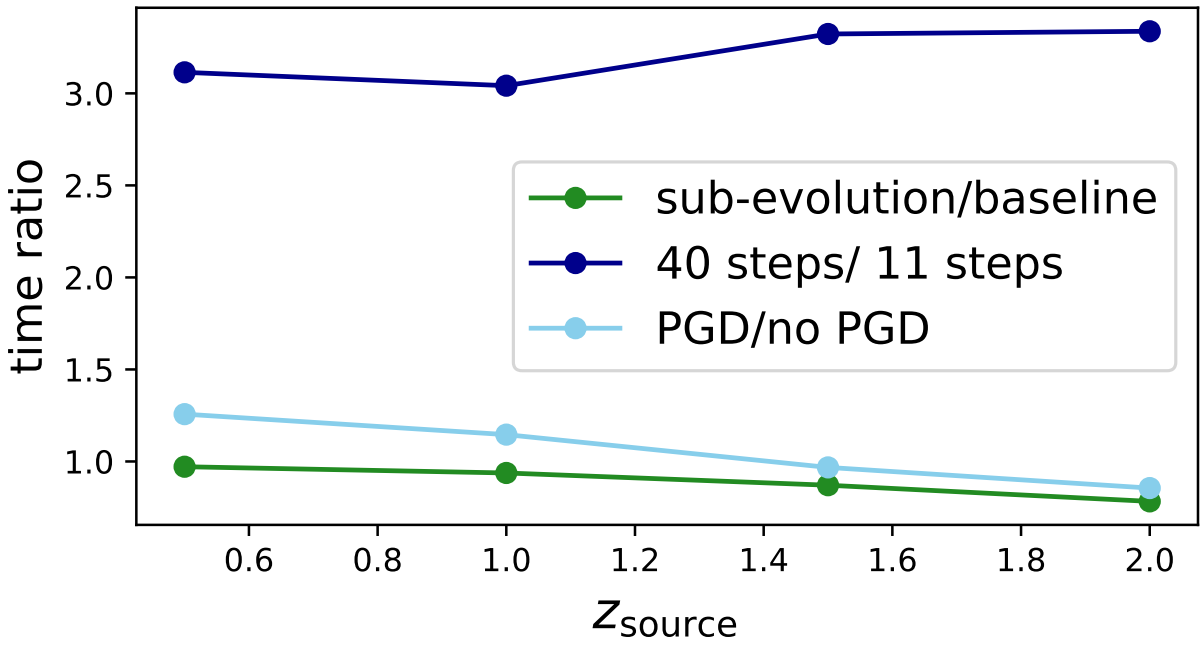



Figure 4.11: A comparison of different settings of the simulation package in terms of computation time. Using an 11-step simulation instead of a 40-step simulation reduces the computation time by more than a factor of 3. This reduction in computation time without paying a significant penalty in simulation accuracy is made possible by the PGD enhancement and the sub-evolution step. Neither of those enhancements change the computations times significantly (green and light blue lines). All simulations in this test were run with $256^3$ particles and a $256$ Mpc/h box.

No. DE-AC02-05CH11231. The authors thank Uroš Seljak for useful discussions and feedback and François Lanusse for testing of the package and helpful feedback.

## 4.6 Appendix A: Differentiability with respect to Cosmological Parameters

Since the forward model itself depends on cosmological parameters through the evolution of particle positions and the angular diameter distance which enters the lensing projection, an accurate inference algorithm needs to take these dependencies into account. To this end MADLens provides the additional functionality of derivatives with respect to the cosmological parameters $\Omega_{m0}$ and $\sigma_8$.

This novel application of derivatives requires both power spectra, particle evolution, and comoving distance calculations written as functions of cosmological parameters. The comoving distance calculation and derivative are trivial, and we use the standard definition [244],

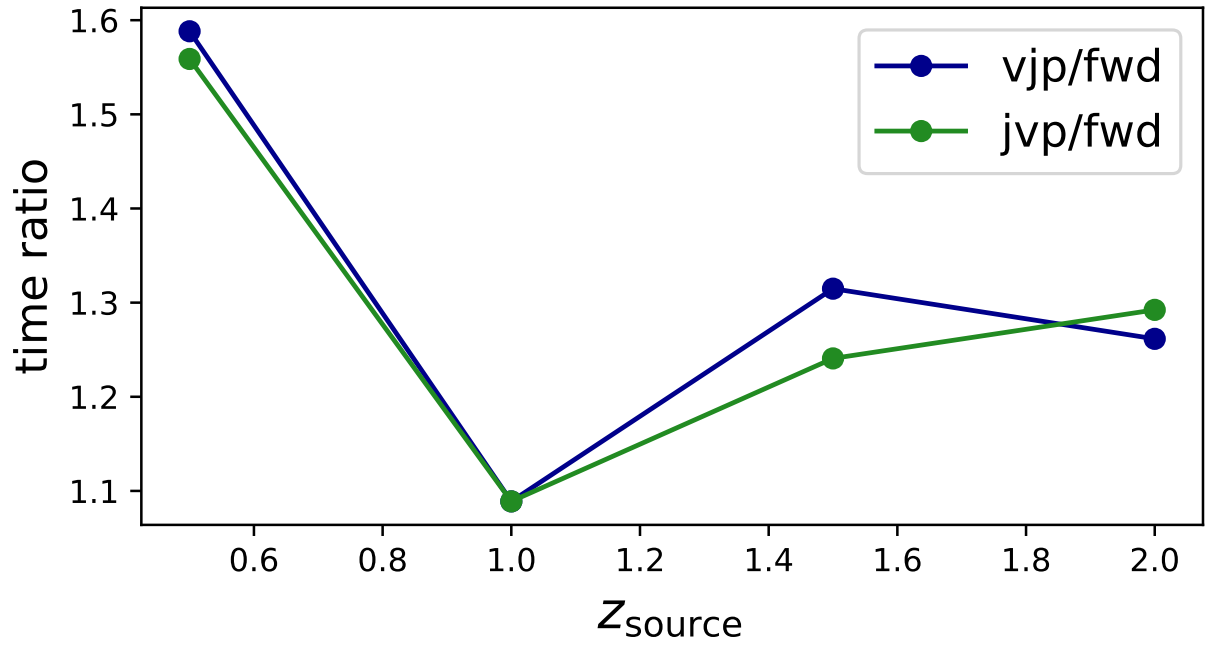



Figure 4.12: Computation time of vector-Jacobian (vJp) and Jacobian-vector product (Jvp) in units of computation time of the forward model for different source redshifts. The vJp and Jvp require only slightly longer computation times than the forward model.

$$\chi = \frac{c}{H_0} \int_0^z \frac{dz'}{E(z')^{1/2}}, \tag{4.7}$$

where $E = \Omega_{m0}(1+z)^3 + \Omega_{k0}(1+z)^2 + \Omega_{\Lambda 0}$. For a flat cosmology, we take $\Omega_{k0} \sim 0$ and $\Omega_{\lambda 0} \sim 1 - \Omega_{m0}$.

For the power spectrum that is applied to the initial density field, we use the transfer function defined in [82] (EH-Transfer) with the inclusion of baryonic acoustic oscillation (BAO) wiggles. This is much simpler and computationally less involved than obtaining gradients of standard Boltzmann solvers with respect to cosmological parameters. Compared to the matter power spectrum obtained from the Boltzmann package, CLASS [26], which is used for cosmological calculation throughout MADLens, we find discrepancies at a maximum of the $\sim 5\%$ level. We find that by using the EH-transfer, we slightly overestimate power on all scales with the largest discrepancies at those corresponding to BAO wiggles.

To show that this overestimation is within reason, we generate multiple convergence maps with both the CLASS and differentiable Eisenstein and Hu power spectrum. We show the absolute difference of the power spectra of these maps in Figure 4.13 and plot the cosmic variance for comparison. We find that the difference lies below the $1\sigma$ limit which we take to imply that our

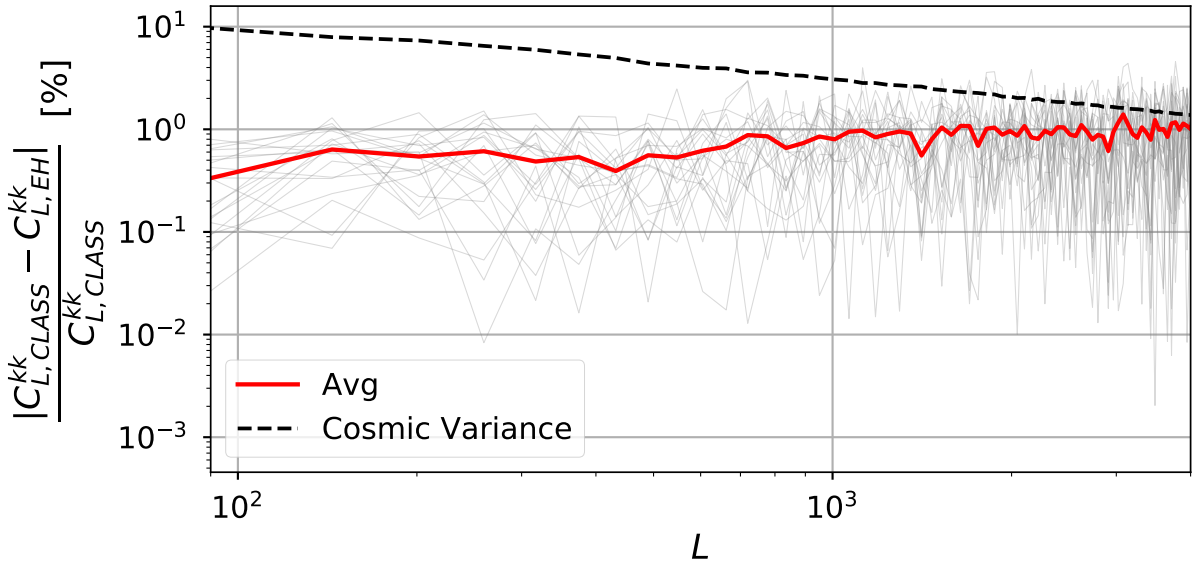



Figure 4.13: We calculate 20 convergence maps using both CLASS and Eisenstein & Hu transfer functions and see a strong agreement in values. The percent error of the power with respect to the CLASS transfer function convergence maps (red line) remains below the cosmic variance (dotted black line) for most modes with negligible deviations above at small scales.

implementation of the EH-transfer function and power describes the initial spectrum well within the required accuracy.

The particle initial conditions and evolution, too, is dependent on the cosmology and we use a finite differencing scheme on the Lagrangian Perturbation Theory initial conditions, as well as the momentum and position updates in FastPM. This allows the computation of derivatives for the initial conditions, kick, and drift factors without the need for analytical solutions to the derivatives with respect to $\Omega_{m0}$. This method works by storing the finite difference of a function on the gradient tape, and caches the cosmology variables to avoid re-computation of the particle mesh. During forward propagation, the function is run as normal, and during back propagation two cached cosmology objects with perturbed parameter values are used to compute the finite difference. This scheme can be made applicable to any function which is not highly sensitive to parameters, and while we use the analytic solution for derivatives such as the EH-transfer function, it should be noted that it is feasible to apply the finite differencing method to linear power and cosmology methods from Boltzmann codes such as CLASS.

Similar to the finite difference test in section 4.4, we test the accuracy of the Jvp for $\sigma_8$. We add and subtract a small offset $\delta$ from $\sigma_8$ ($\delta = 10^{-10}$), compute convergence maps from both of these configurations and take their difference. We then compare this to the Jvp at the central $\sigma_8$ value with a vector of $2\delta$ to multiply the Jacobian. The resulting histograms are shown in Figure 4.14. The derivative of $\sigma_8$ is in excellent agreement with the finite differencing result.

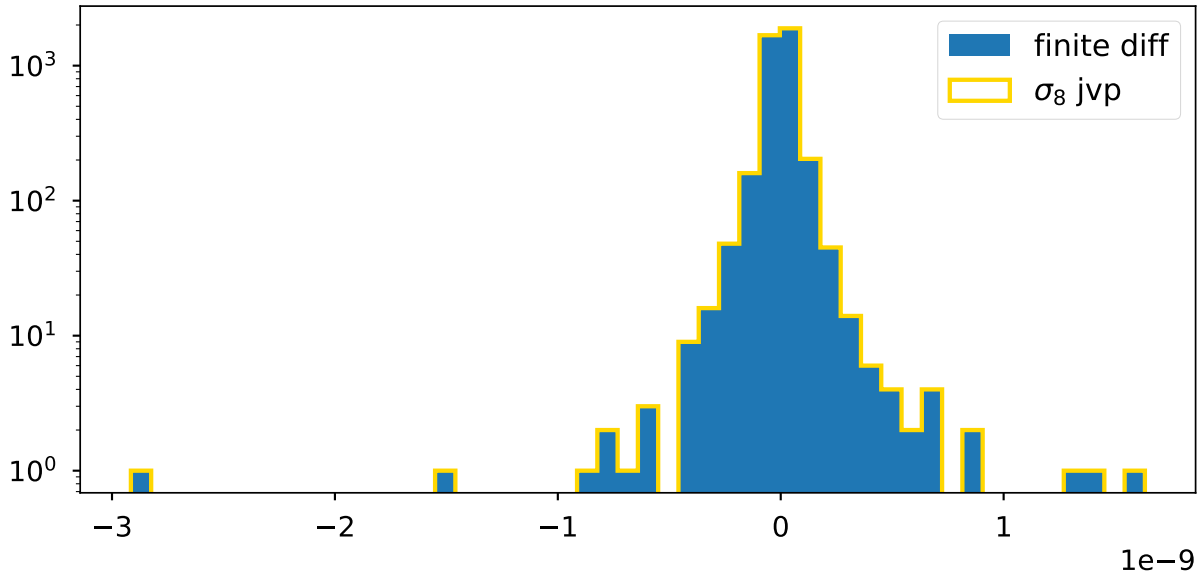



Figure 4.14: We show that the automatic derivative and finite differencing agree well for $\sigma_8$. We used a small offset $\delta = 10^{-10}$ and histogram the finite-differenced convergence maps (*blue*) and Jvp outputs (*gold*)

While $\sigma_8^2$ only enters linearly, the model's dependence on $\Omega_{m0}$ is more complicated. The derivative with respect to $\Omega_{m0}$ therefore requires a more in depth testing to ensure accurate derivatives are being taken. We test the Jacobian against finite differencing results for multiple modes individually. In Figure 4.15, we show the results of cross-correlating the finite difference results with the Jvp outputs and verify the accuracy of the automatic derivative.

As a check for the vJp against $\Omega_{m0}$, we construct a scalar by computing the finite differencing of the sum of the squared convergence maps and ensure that this value is equal to the vJp when the central convergence map is used as the vector in automatic differentiation.

$$\frac{\sum_{i=0}^{N} \kappa_{i,+\delta}^2 - \sum_{i=0}^{N} \kappa_{i,-\delta}^2}{2\delta} = 2vjp_{\Omega_{m0}}(v = \kappa) \tag{4.8}$$

We find that the values agree at the $\sim 5\%$ level irrespective of the choice of $\delta$.

## 4.7 Appendix B: Details about FastPM and the sub-evolution scheme

The FastPM evolution equations are based on the Zeldovich approximation. A FastPM step from scale factor $a_i$ to $a_f$ starts by calculating the gravitational potential from the current particle positions. This step involves Fourier transforms and is the computational bottleneck of the algorithm.

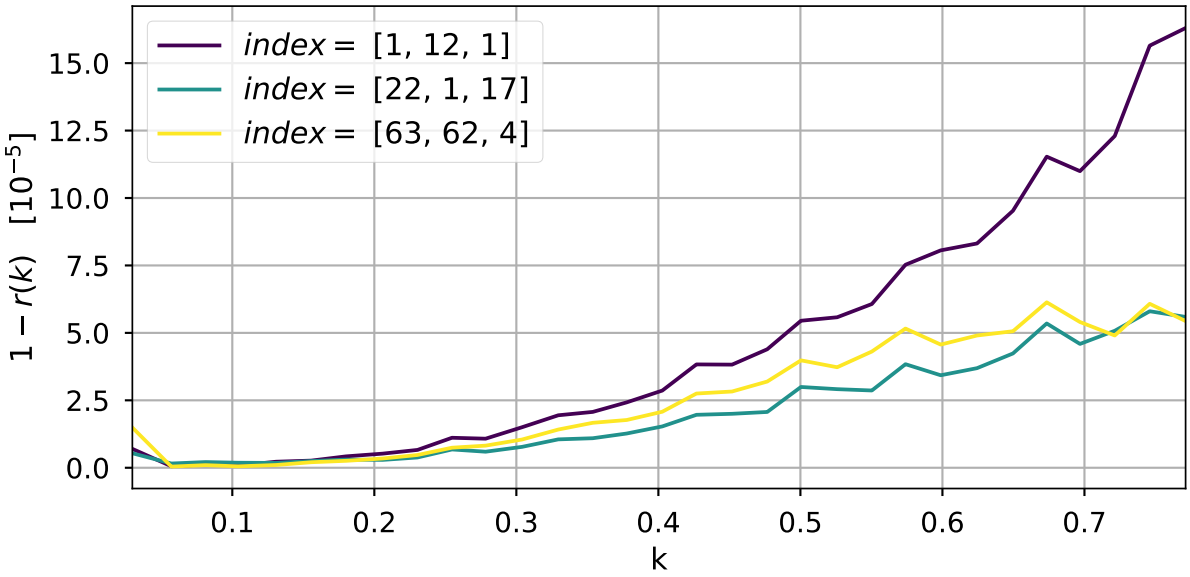



Figure 4.15: On each scale, the finite difference and automatic derivative of $\Omega_{m0}$ agree to high accuracy. We show this by randomly selecting 3 modes to excite individually while setting all other modes to zero. In this plot, each line corresponds to the cross correlation between convergence maps generated when each of these selected modes is set to one, while the rest of the field is set to 0 for a $64^3$ mesh, and the corresponding Jvp. The indices used are specified in the legend. We find that irrespective of index choice, these correlations agree at the $10^{-5}$ level.

The gravitational force is then used to update the particle momenta. The particle positions $x(a_0)$ are updated by the increment

$$\Delta x = p(a_c)\mathcal{D}(a_i, a_c, a_f), \tag{4.9}$$

where $p(a_c)$ are the particle momenta at the central scale factor $a_c = \sqrt{a_i a_f}$, and $\mathcal{D}(a_i, a_c, a_f)$ is the drift factor that approximates near $a_c$ the integral

$$\int_{a_f}^{a_i} \frac{\mathrm{d}a}{a^3 E(a)}. \tag{4.10}$$

For more details on the algorithm, we refer the reader to [89]. In the sub-evolution scheme, the particle positions update is split into two steps. We start by taking half a step,

$$\Delta x_{1/2} = p(a_c)\mathcal{D}(a_i, a_c, a_c), \tag{4.11}$$

and pass these positions to the projection function, where they are further evolved (in either direction) according to the individual particle scale factors $a_{\text{particle}}$,

$$\Delta x_{\text{interp}} = p(a_c)\mathcal{D}(a_{\text{particle}}, a_c, a_c), \tag{4.12}$$

before being projected onto the source planes. This update does not get saved, which is why another step by

$$\Delta x_{1/2} = p(a_c)\mathcal{D}(a_c, a_c, a_f), \tag{4.13}$$

reaches the final positions of a full FastPM step. Taking these additional steps does not require additional Fourier transforms, and hence does not add significantly to the computational complexity of the simulation.

# Chapter 5

# Learning effective physical laws for generating cosmological hydrodynamics with Lagrangian Deep Learning

The contents of this chapter were originally published in [54],

Learning effective physical laws for generating cosmological hydrodynamics with Lagrangian Deep Learning
*Dai B., Seljak U.*(arXiv:2010.02926) PNAS 118.16 (2021)

The goal of generative models is to learn the intricate relations between the data to create new simulated data, but current approaches fail in very high dimensions. When the true data-generating process is based on physical processes, these impose symmetries and constraints, and the generative model can be created by learning an effective description of the underlying physics, which enables scaling of the generative model to very high dimensions. In this work we propose Lagrangian Deep Learning (LDL) for this purpose, applying it to learn outputs of cosmological hydrodynamical simulations. The model uses layers of Lagrangian displacements of particles describing the observables to learn the effective physical laws. The displacements are modeled as the gradient of an effective potential, which explicitly satisfies the translational and rotational invariance. The total number of learned parameters is only of order 10, and they can be viewed as effective theory parameters. We combine the N-body solver FastPM with LDL and apply them to a wide range of cosmological outputs, from the dark matter to the stellar maps, gas density and temperature. The computational cost of LDL is nearly four orders of magnitude lower than the full hydrodynamical simulations, yet it outperforms it at the same resolution. We achieve this with only of order 10 layers from the initial conditions to the final output, in contrast to typical cosmological simulations with thousands of time steps. This opens up the possibility of analyzing cosmological observations entirely within this framework, without the need for large dark-matter simulations.

## 5.1 Introduction

Numerical simulations of large scale structure formation in the universe are essential for extracting cosmological information from the observations [81, 141, 298, 66, 9, 303, 63]. In principle hydrodynamical simulations are capable of predicting the distribution of all observables in the universe, and thus can model observations directly. However, running high resolution hydrodynamical simulations at volumes comparable to the current and future sky surveys is currently not feasible, due to its high computational costs. The most widely used method is running gravity-only N-body simulations, and then populating baryons in the halo catalogs with semi-analytical approaches such as halo occupation distribution (HOD) [21], or halo assembly models [18]. However, these methods make strong assumptions such as the halo mass being the main quantity controlling the baryonic properties. In addition, many of the cosmological observations such as X-ray emission and Sunyaev-Zeldovich emission are based on hydrodynamic gas properties such as gas density, temperature, pressure etc., which cannot be modeled in the dark matter only simulations.

Deep learning methods provide an alternative way to model the cosmological observables. A number of papers view the task as an image-to-image translation problem, i.e., they take in pixelized matter density field as input data, and output the target pixelized observable field. These methods either model the conditional probability distribution $p(y_{target}|x_{input})$ using deep generative models such as GANs [107] and VAEs [162, 269], or learn a mapping $x_{input} \mapsto y_{target}$ with deep convolutional neural networks (DCNN). Previous work in this area covers a wide array of tasks, such as identifying halos (protohalos) [224, 20, 264, 22], producing 3D galaxy distribution [364], generating tSZ signals [321], predicting dark matter annihilation feedback [188], learning neutrino effects [105], emulating high resolution features from low resolution simulations [171, 184] etc.

Unlike these methods that work in pixel (Eulerian) space and treat the field as images, another way to model the dynamics is to adopt the Lagrangian scheme, i.e., trace the motion of individual particles or fluid elements by modeling their displacement field. The displacement field contains more information than the density field, as different displacement fields can produce the same density field, and is in general more Gaussian and linear than the density field. Existing methods in this space only cover the dark matter, e.g. approximate N-body solvers [312, 89] and DCNN [124].

In this work we propose a novel deep learning architecture, Lagrangian Deep Learning (LDL), for modeling both the cosmological dark matter and the hydrodynamics, using the Lagrangian approach. The model is motivated by the effective theory ideas in physics, where one describes the true process, which may be too complicated to model, with an effective, often coarse grained, description of physics. A typical example is the effective field theory, where perturbative field theory is supplemented with an effective field theory terms that obey the symmetries, and are an effective coarse-grained description of the non-perturbative small scale effects. The resulting effective description has a similar structure as the true physics, but with free coefficients that one must fit for, and that account for the non-perturbative small scale effects.

## 5.2 Lagrangian Deep Learning

Cosmological dark matter and baryon evolution can be described by a system of partial differential equations (PDE) coupling gravity, hydrodynamics, and various sub-grid physics modeling processes such as the star formation, which are evolved in time from the beginning of the universe until today. One would like to simulate a significant fraction of the observable universe, while also capturing important physical processes on orders of magnitude smaller scales, all in three dimensions. As a result the resulting dynamical range is excessive even for the modern computational platforms. As an example, the state of the art Illustris TNG300-1 [252, 228, 208, 302, 232] has of order $10^{10}$ particles, yet simulates only a very small fraction of the observable universe. The lower resolution TNG300-3 reduces the number of particles by 64, at a cost of significantly reducing the realism of the simulation.

An effective physics approach is to rewrite the full problem into a large scale problem that we can solve, together with an effective description of the small scales which we cannot resolve. In theoretical physics this is typically done by rewriting the Lagrangian such that it takes the most general form that satisfy the symmetries of the problem, with free coefficients describing the effect of the small scale coarse-graining. In cosmology the large scale evolution is governed by gravity, which can easily be solved perturbatively or numerically. Effective descriptions using perturbative expansions exist [39], but fail to model small scales and complicated baryonic processes at the map level. While spatial coarse graining is the most popular implementation of this idea, one can also apply it to temporal coarse graining as well. A typical PDE solver requires many time steps, which is expensive. Temporal coarse graining replaces this with fewer integration time steps, at a price of replacing the true physics equations with their effective description, while ensuring the true solution on large scales, where the solution is known [312, 89].

Here we take this effective physics description idea and combine it with the deep learning paradigm, where one maps the data through several layers consisting of simple operations, and trains the coefficients of these layers on some loss function of choice. While machine learning layers are described with neural networks with a very large number of coefficients, here we will view a single layer as a single time step PDE solver, using a similar structure as the true physical laws. [1] This has the advantage that it can preserve the symmetries inherent in the problem. The main symmetry we wish to preserve in a cosmological setting is the translational and rotational symmetry: the physical laws have no preferred position or direction. But we also wish to satisfy the existing conservation laws, such as the dark matter and baryon mass conservation.

A very simple implementation of these two requirements is Lagrangian displacements of particles describing the dark matter or baryons. We displace the particles using the gradient of a potential, and mass conservation is ensured since we only move the particles around. To ensure the translation and rotation symmetry within the effective description we shape the potential

[1] In the rest of this paper we will refer to each step of the model as one layer, to emphasize the fact that our model is parametrized by free parameters that need to be trained on hydro simulations. However, one should note that in this research we do not use any actual neural networks. Since the FastPM time step can be viewed as a special case of our model, we will also refer to each FastPM time step as one layer, even though the parameters (e.g. kick and drift factors) are determined by physics rather than fitting to some loss functions.

in Fourier space, such that it only depends on the amplitude of the Fourier wave vector. The potential gradient can be viewed as a force acting upon their acceleration via the Newton's law, and the shaping of the potential is equivalent to the radial dependence of the force. This description requires particle positions and velocities, so it is a second order PDE in time. We will use this description for the dark matter. However, for baryons we can simplify the modeling by assuming their velocity is the same as that of the dark matter, since velocity is dominated by large scales where the two trace each other. In this case we can use the potential gradient to displace particle positions directly, so the description becomes effectively first order in time. Moreover, by a simple extension of the model we can apply this concept to the baryonic observables such as the gas pressure and temperature, where conservation laws no longer apply. A complete description also requires us to define the source for the potential. In physics this is typically some property of the particles, such as mass or charge. Here we wish to describe the complicated nonlinear processes of subgrid physics, as well as coarse graining in space and time. Motivated by gravity we will make the simplest possible assumption of the source being a simple power law of the density, using a learned Green's function to convert to the potential. Since we wish to model several different physics processes we stack it into multiple layers. Because the model takes in the particle data and models the displacement field from the Lagrangian approach using multiple layers, we call this model Lagrangian Deep Learning (LDL).

Our specific goal is to model the distribution of dark matter and hydrodynamic observables starting from the initial conditions as set in the early universe, using an effective description that captures the physics symmetries and conservation laws. An example of such a process applied to time and spatial coarse graining is the dark matter evolution with a few time steps only, which combines ideas such as the approximate N-body solvers, with a force sharpening process called the Potential Gradient Descent (PGD) to capture the coarse graining [53, 58]. We first use FastPM [89], a quasi particle-mesh (PM) N-body solver, which ensures the correct large scale growth at any number of time steps, since the kick and drift factors of the leapfrog integrator in FastPM are modified following the linear (Zel'dovich) equation of motion. FastPM has a few layers only (typically 5-10) and uses particle displacements. It is supplemented by one additional layer of PGD applied to position only to improve the dark matter distribution on small scales. All of the steps of this process are in the LDL form, so can be viewed as its initial layers. The resulting dark matter maps are shown in figure 5.1 and show an excellent agreement with the full N-body simulation of Illustris TNG, which is also confirmed by numerical comparisons presented in [53]. This application is not learning new physics, but is learning the effective physics description of both time and spatial coarse graining: instead of 1000+ time steps in a standard N-body simulation we use only 10, and instead of the full spatial resolution we will use a factor of 64 reduced mass resolution.

Here we wish to extend these ideas to the more complex and expensive problem of cosmological hydrodynamics, where we wish to learn its physics using an effective description. Baryons are dissipative and collisional, with many physical processes, such as cooling, radiation, star formation, gas shocks, turbulence etc. happening inside the highest density regions called dark matter halos. One can add displacements to the dark matter particles to simulate these hydrodynamic processes, such that the particles after the displacement have a similar distribution as the baryons. Enthalpy

Gradient Descent (EGD) is an example of this idea [53]: one adds small scale displacement to the dark matter particles to improve the small-scales of the low resolution approximate simulations, and to model the baryonic feedback on the total matter distribution. Motivated by these methods, we propose to model this displacement field by

$$\mathbf{S} = \alpha \nabla \hat{\mathbf{O}}_{\mathbf{G}} f(\delta), \tag{5.1}$$

where $\alpha$ is a learnable parameter, $\delta$ is the matter overdensity as output by the initial layers (FastPM and LDL on dark matter layer), $f(\delta)$ is the source term and can be an arbitrary function of $\delta$. Here we choose it to be a power law

$$f(\delta) = (1+\delta)^{\gamma}, \tag{5.2}$$

with $\gamma$ a learnable parameter. $\hat{\mathbf{O}}_{\mathbf{G}}$ is the Green's operator, and can be written explicitly as

$$\hat{\mathbf{O}}_{\mathbf{G}} f(\delta) = \int G(\mathbf{x}-\mathbf{x}') f(\delta(\mathbf{x}')) d\mathbf{x}', \tag{5.3}$$

where $G(\mathbf{x}-\mathbf{x}')$ is the Green's function and we have used $G(\mathbf{x},\mathbf{x}') = G(\mathbf{x}-\mathbf{x}')$ due to translational symmetry. The convolution in above equation can be easily calculated in Fourier space as $\hat{\mathbf{O}}_{\mathbf{G}} f(\delta) = G(\mathbf{k}) f(\delta)$, and we further have $G(\mathbf{k}) = G(k)$ because of the rotational symmetry of the system. Following the PGD model, we model $\hat{\mathbf{O}}_{\mathbf{G}}$ in Fourier space as

$$\hat{\mathbf{O}}_{\mathbf{G}} = \exp(-(k_h/k)^2) \exp(-(k/k_l)^2)\, k^n, \tag{5.4}$$

where $k_h$, $k_l$ and $n$ are additional learnable parameters. The high pass filter $\exp(-(k_h/k)^2)$ prevents the large scale growth, since the baryonic physics that we are trying to model is an effective description of the small scale physics, while the large scales are correctly described by the linear perturbative solution enforced by FastPM. Together with the low pass filter $\exp(-(k/k_l)^2)$, which has the typical effective theory form, the operator $\hat{\mathbf{O}}_{\mathbf{G}}$ is capable of learning the characteristic scale of the physics we are trying to model. Note that both the source $f(\delta)$ and the shape of $\hat{\mathbf{O}}_{\mathbf{G}}$ characterize the complex baryon and subgrid physics and cannot be derived from first principles. They can only be learned from high resolution hydrodynamical simulations. As a special case, Equation 5.1 can be used to represent the gravitational force if we set $\alpha = 4\pi G\bar{\rho}$, $\gamma = 1$, $k_h = 0$, $k_l = \infty$ and $n = -2$, but we allow all these parameters to vary in order to model the physics that is different from gravity. Equation 5.1 - 5.4 defines the displacement field of one Lagrangian layer. We can stack multiple such layers to form a deep learning model, where each layer takes the particle output from the previous layer (which determines $\delta$ in Equation 5.1) and adds additional displacements to the particles. Such a deep model will be able to learn more complex physics. The idea is that different layers can focus on different physics components, which will differ in terms of the scale dependence of the potential and its gradient, as well as in terms of the source density dependence.

We have made the assumption that the displacement field can be approximated by a sequence of irrotational vector fields, which are modeled by the gradient of some effective potential fields. This parameterization, though motivated by physics, is not derived from first principles and its

effectiveness has to be examined by experiments. In principle, one could also use second order equations for modeling the displacement fields, just like the true dynamics, at the cost of introducing one more free parameter per layer. Note that here we are only trying to effectively mimic the missing baryonic physics rather than actually simulating it, so the order of the equation does not matter. As we show in this paper, the first order equation works well enough for the resolution we considered, and thus we do not consider the second order parametrizations.

The final output layer is modeled as a nonlinear transformation on the particle density field:

$$F(x) = \mathrm{ReLU}(b_1(1 + \delta'(x))^\mu - b_0), \tag{5.5}$$

where $F$ is the output target field, $\mathrm{ReLU}(x)$ is the rectified linear unit, which is zero if $x < 0$ and $x$ otherwise, $\delta'$ is the particle overdensity field after the displacement, and $b_0$, $b_1$ and $\mu$ are learnable parameters. This is motivated by the physics processes that cannot be modeled as a matter transport (i.e. displacement). In the example of stars, the Lagrangian displacement layers are designed to learn the effect of gas cooling and collapse. After these displacement layers, the particles are moved towards the halo center, where protogalaxies are formed and we expect star formation to happen in these dense regions. This star formation process will be modeled by Equation 5.5: the $\mathrm{ReLU}$ thresholding aims at selecting the high density regions where the star formation happens. Such thresholding is typical of a subgrid physics model: in the absence of this thresholding we would need to transport all of the particles out of the low density regions where the star formation does not happen, a process that does not have a corresponding physical model.

The baryon process in the universe also leaves imprints on the total matter distribution [275, 326, 138]. This feedback is crucial for accurately predicting the total matter distribution (e.g. in weak lensing applications). In this research, however, we mainly focus on modeling the distribution of baryon tracers and do not consider this feedback effect. We would like to refer interested readers to the Enthalpy Gradient Descent (EGD) model [53], which is a special case of LDL and has been shown to successfully model the baryonic feedback on the total matter distribution. We expect that LDL can also be applied to model this effect in a similar manner.

In this work we use both FastPM and N-body simulations, combining them with LDL to predict the baryon observables from the linear density map. We consider modeling the stellar mass, kSZ signal, tSZ signal and X-ray at redshift $z = 1$, $z = 0.5$ and $z = 0$. The dark matter particles are firstly evolved to these redshifts with FastPM, and then passed to the LDL networks for modeling the baryons. The parameters in LDL are optimized by matching the output with the target fields from TNG300-1 hydrodynamical simulation [252, 228, 208, 302, 232]. Since the kSZ signal is proportional to the electron momentum $n_e v_z$, the tSZ signal is proportional to the electron pressure $n_e T$, and the X-ray emissivity is approximately proportional to $n_e^2 T^{0.5}$ (we only consider the bremsstrahlung effect and ignore the Gaunt factor), we will model these fields in the rest of this paper. See the Materials and Methods Section at the end of the paper for details on the FastPM/N-body + LDL hybrid simulations, as well as the training of the LDL parameters.

Apart from FastPM, we also consider combining LDL models with full N-body simulations. We take the particle data at redshift $z = 1$, $z = 0.5$ and $z = 0$ from TNG300-3-Dark, a low resolution dark-matter-only run of the TNG300 series, and feed the particles to LDL models. In the

Table 5.1: The numerical parameters of LDL hybrid simulations, low resolution TNG300-3 and the target TNG300-1 hydrodynamical simulations

| | FastPM + LDL | TNG300-3-Dark + LDL | TNG300-3 | TNG300-1 |
|---|---|---|---|---|
| $N_{\rm particle}$ | $625^3$ | $625^3$ | $N_{\rm DM} = 625^3$<br>$N_{\rm gas} = 625^3$ | $N_{\rm DM} = 2500^3$<br>$N_{\rm gas} = 2500^3$ |
| Force / Mesh Resolution ($h^{-1}$ckpc) | 164 (FastPM)<br>328 (LDL) | 4.0 (TNG-Dark)<br>328 (LDL) | $\epsilon_{\rm DM,*} = 4.0$<br>$\epsilon_{\rm gas} = 1.0$ | $\epsilon_{\rm DM,*} = 1.0$<br>$\epsilon_{\rm gas} = 0.25$ |
| Number of Time Steps / layers | $N_{\rm FastPM} = 10$<br>$N_{\rm LDL,*} = 4$ | $N_{\rm TNG} = 9201$<br>$N_{\rm LDL,*} = 3$ | $209,161$ | $6,203,062$ |
| CPU Time | $T_{\rm IC} = 2.3$ h<br>$T_{\rm FastPM} = 5.1$ h<br>$T_{\rm LDL,*} = 0.4$ h | $T_{\rm TNG} = 5.9$ kh<br>$T_{\rm LDL,*} = 0.3$ h | 0.05 Mh | 34.9 Mh |

The LDL parameters for generating stellar mass. The architecture for other observables can be found in Table 5.2. The total CPU time for LDL is 7.8 hours, compared to $5 \times 10^4$ for the full hydro TNG300-3. Despite this the LDL outperforms the full hydro at the same resolution in all of the outputs. In this paper we are primarily concerned with a proof of principle and both FastPM and LDL are run with Python. We expect the CPU time to be further reduced if running them with C.

next section we will compare the performance of these two hybrid simulations against the target high resolution hydrodynamical simulation.

We summarize the numerical parameters of these simulations in Table 5.1. We also list TNG300-3, the low resolution hydrodynamic run of TNG300. TNG300-3 has the resolution of our hybrid simulations, and is a natural reference to compare the performance of our models with. Note that the mass resolution, force / mesh resolution and time resolution of our hybrid simulations are significantly lower than the target simulation, and the N-body simulation and deep learning networks are also much cheaper to run compared to simulating hydrodynamics. As a result, the FastPM-based and N-body-based hybrid simulations are 7 and 4 orders of magnitude cheaper than the target simulation, respectively. When comparing to TNG300-3, our hybrid simulations are still 4 and 1 orders of magnitude cheaper, respectively, and we show that by being trained on the high resolution TNG300-1 our simulations are superior to TNG300-3, and comparable to TNG300-1.

## 5.3 Results

We show in Figure 5.1 the visualization of slices of the input linear density field and the output dark matter of our FastPM-based hybrid simulation, as well as the target fields in hydrodynamical simulation. Visual agreement between the two is very good. The results are shown for the dark matter density, stellar mass density, electron momentum density $n_e v_z$, where $n_e$ is electron density

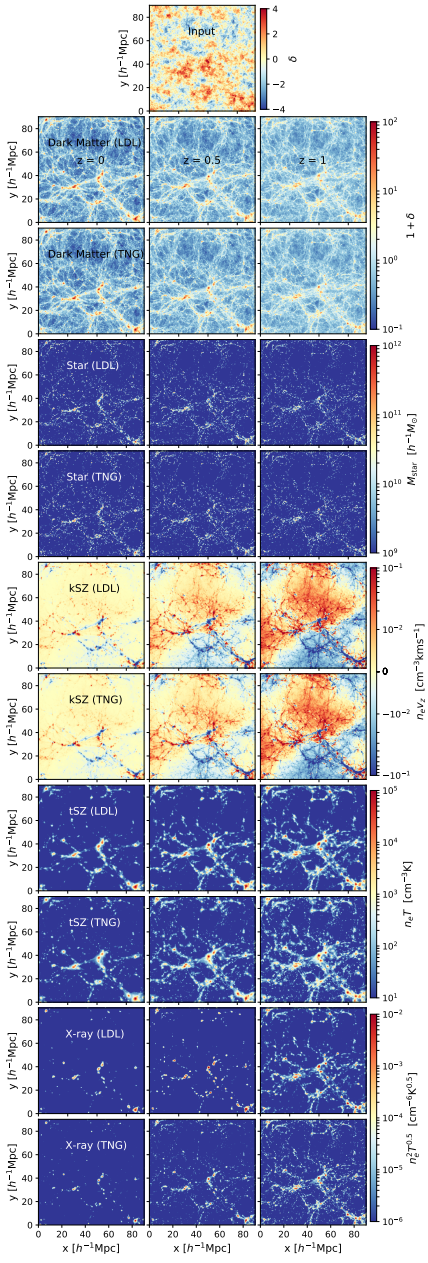



Figure 5.1: Visualization of slices of the simulations: the first row is the input linear density field. The 2nd, 4th, 6th, 8th, and 10th rows are predictions of dark matter overdensity, stellar mass, electron momentum density $n_e v_z$ (kSZ signal), electron pressure $n_e T$ (tSZ signal) and $n_e^2 T^{0.5}$ (X-ray signal) from our FastPM+LDL hybrid simulation, respectively. The 3rd, 5th, 7th, 9th, and 11th rows are the corresponding target fields from the TNG300-1 hydrodynamical simulation. The left panel, middle panel and right panel are from redshift $z = 0$, $z = 0.5$, and $z = 1$, respectively. The slices are from a $90.2 \times 90.2 \times 32.8 h^{-1}\mathrm{Mpc}$ sub-box of the test set.

and $v_z$ radial velocity, electron pressure $n_e T$, where $T$ is the gas temperature, and X-ray emission proportional to $n_e^2 T^{0.5}$.

## Power Spectrum

We measure the summary statistics of these fields and compare them quantitatively. We firstly compare the power spectrum, the most widely used summary statistics in cosmology. We define the transfer function as:

$$T(k) = \sqrt{\frac{P_{\rm predict}(k)}{P_{\rm target}(k)}}, \tag{5.6}$$

and the cross correlation coefficient as:

$$r(k) = \frac{P_{\rm predict,target}(k)}{\sqrt{P_{\rm predict}(k) P_{\rm target}(k)}}, \tag{5.7}$$

where $P_{\rm predict,target}(k)$ is the cross power spectrum between the predicted field and the target field. We show the 3D or 2D power spectrum, transfer function and cross correlation coefficient of the stellar mass overdensity $\delta_*$, electron momentum $n_e v_z$, electron pressure $n_e T$ and X-ray intensity $n_e^2 T^{0.5}$ in Figures 5.2, 5.3, 5.4 and 5.5, respectively. On large scale and intermediate scale our hybrid simulations in general match well with the target fields (except for the case of X-ray, where FastPM-based hybrid simulations fail to predict correct large scale power at redshift 0.5 and 1), while TNG300-3 agreement is generally worse, especially for the stellar mass. The large bias of TNG300-3 stellar mass might be partially due to the fact that the low resolution TNG300-3 cannot resolve the stars in small halos. In contrast, by training on high resolution hydro simulations TNG300-1, our low resolution hybrid simulations are able to model those small galaxies better than the full hydro simulation at the same resolution.

On the small scales all of the predicted fields show some deviations from the targets. We discuss possible reasons for these in the next Section. We also see that the full-N-body-based hybrid simulation normally predicts larger small scale power than the FastPM-based simulation. This is likely due to the fact that the 10-layer FastPM cannot fully model the small halos and halo internal structures, and its simulated dark matter distribution is less clustered on small scale compared to full N-body simulations, making its predicted baryon fields less clustered. Overall, the predicted power spectrum from the N-body-based hybrid simulation is better, although it can predict too much small scale power (e.g. the kSZ signal at redshift 1). This inaccuracy of the small scale matter distribution of FastPM could also explain the inconsistency of its predicted large scale power of X-ray at redshift 0.5 and 1. Since the X-ray signal is proportional to $n_e^2$, its large scale power is correlated to the small scale fluctuations of $n_e$ field. This modeling task is not easy for FastPM, but N-body-based hybrid simulation is able to model it pretty well (Figure 5.5).

The cross correlation coefficients are also shown in these Figures. We observe that the hybrid simulations are significantly better than those of TNG300-3, with the N-body-based hybrid simulation a bit higher than the FastPM-based simulation. Note that in principle the cross correlation

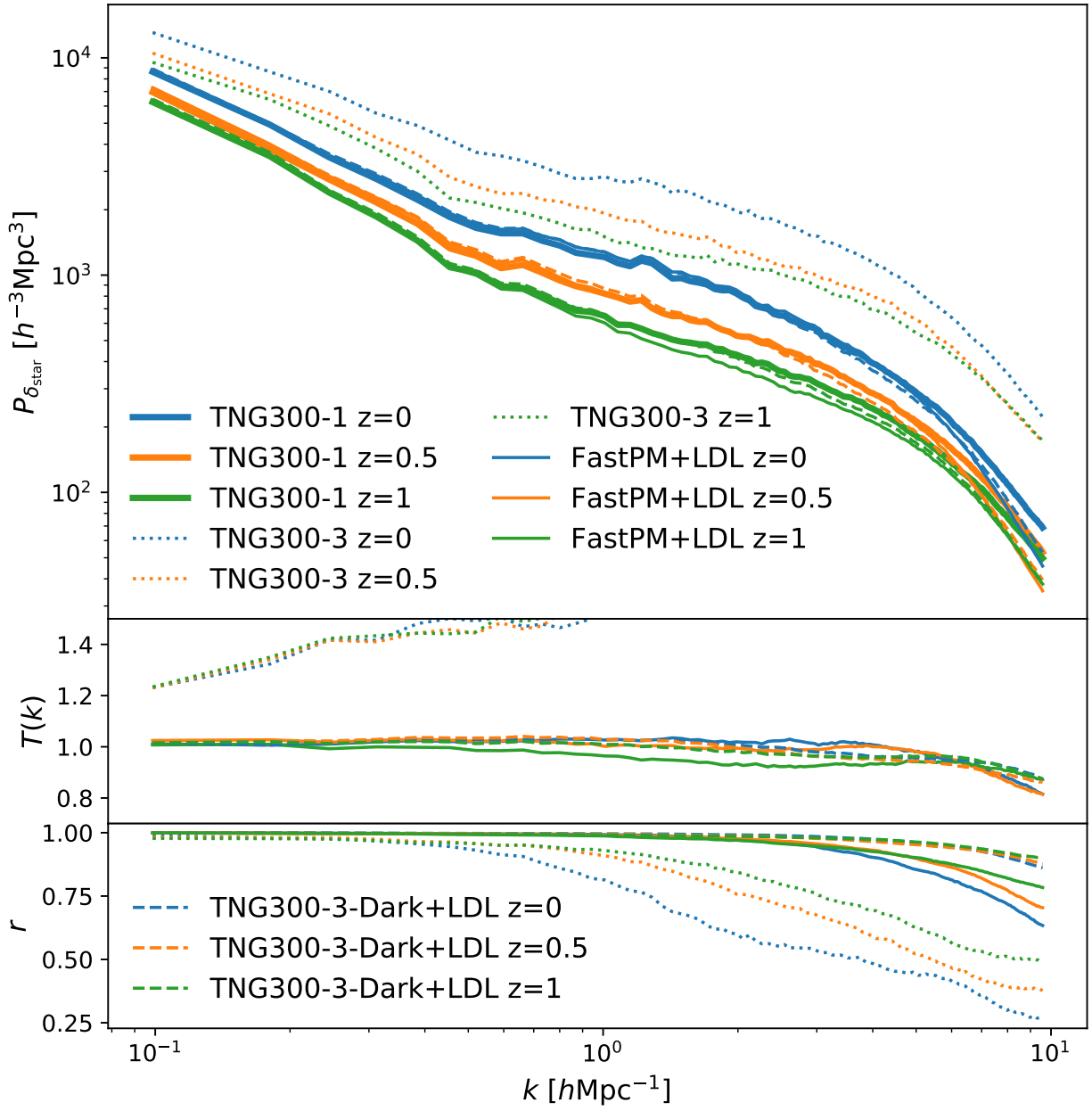



Figure 5.2: Comparison of the test set 3D power spectrum (top panel), transfer function (middle panel) and cross correlation coefficient (bottom panel) of the stellar mass overdensity. We compare LDL hybrid simulations, TNG300-3 and the target TNG300-1 hydrodynamical simulation, at redshifts 0, 0.05 and 1.

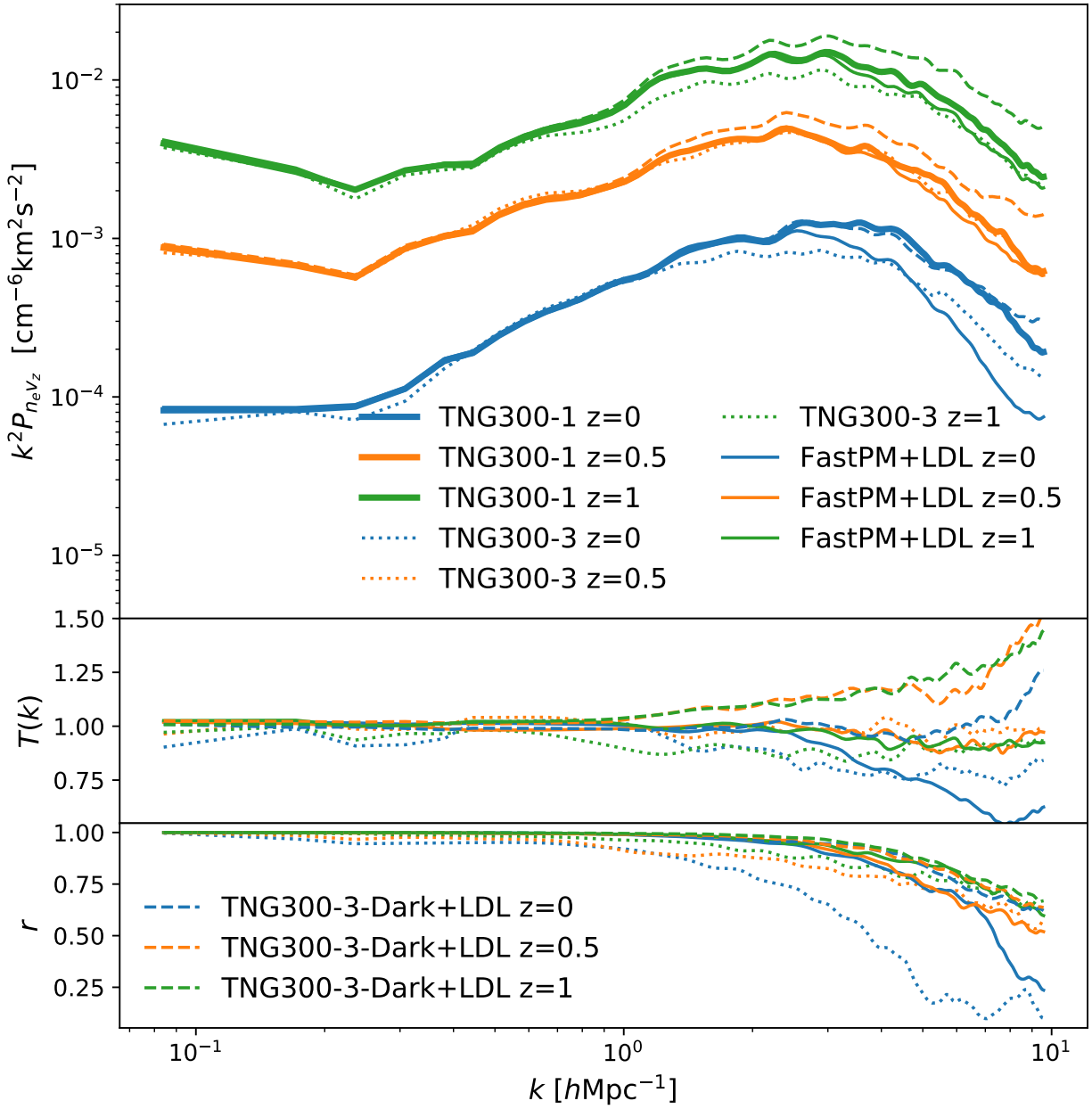



Figure 5.3: Comparison of the test set 2D power spectrum (top panel), transfer function (middle panel) and cross correlation coefficient (bottom panel) of the electron momentum density $n_e v_z$ (proportional to kSZ signal) between the LDL hybrid simulations, TNG300-3 and the target TNG300-1 hydrodynamical simulation. The $90.2h^{-1}\mathrm{Mpc}$ sub-box of the test set is projected along the z-axis for the calculation of 2D power spectrum.

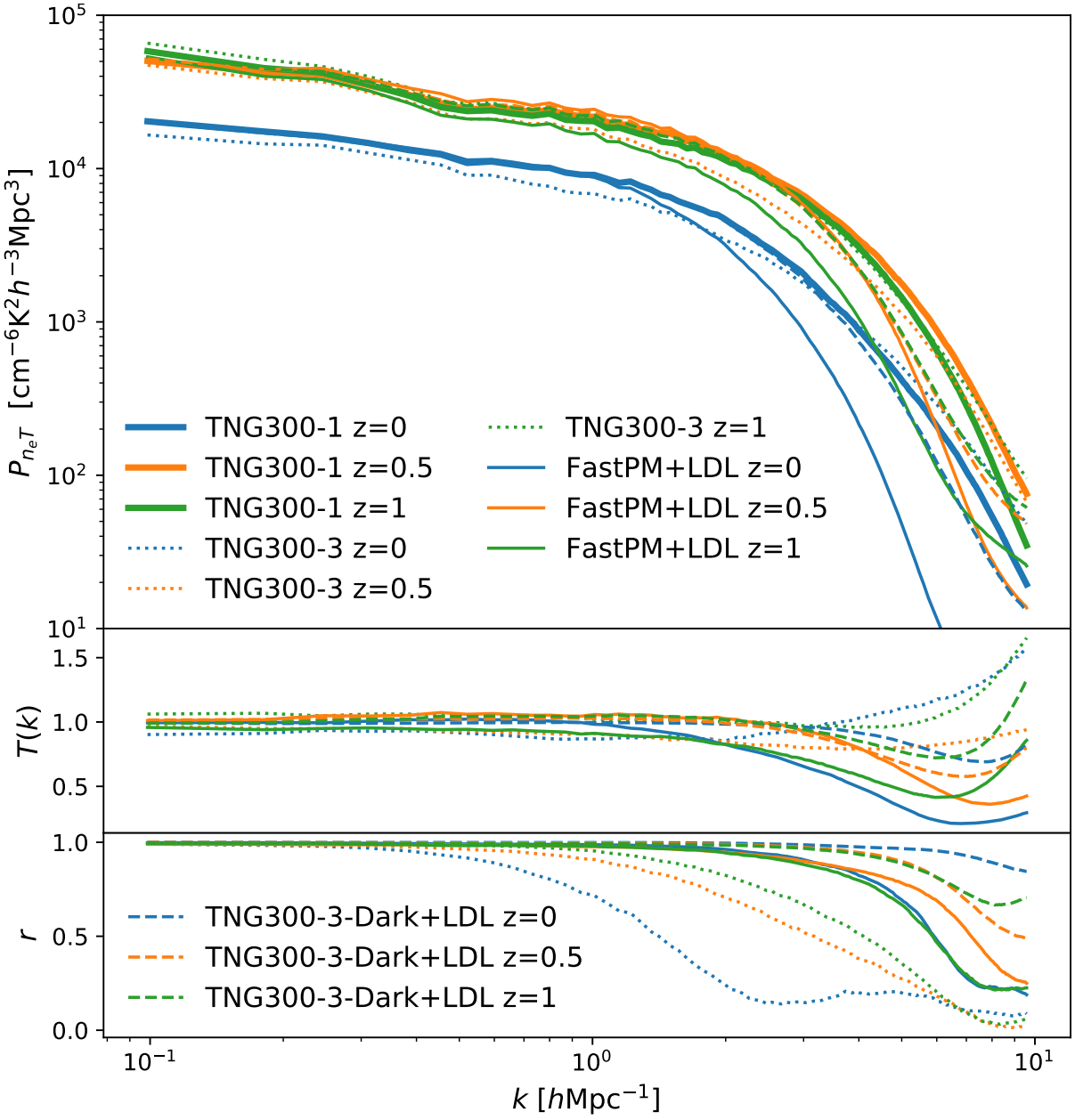



Figure 5.4: Comparison of the test set 3D power spectrum (top panel), transfer function (middle panel) and cross correlation coefficient (bottom panel) of the electron pressure $n_eT$ (proportional to tSZ signal) between the LDL hybrid simulations, TNG300-3 and the target TNG300-1 hydrodynamical simulation.

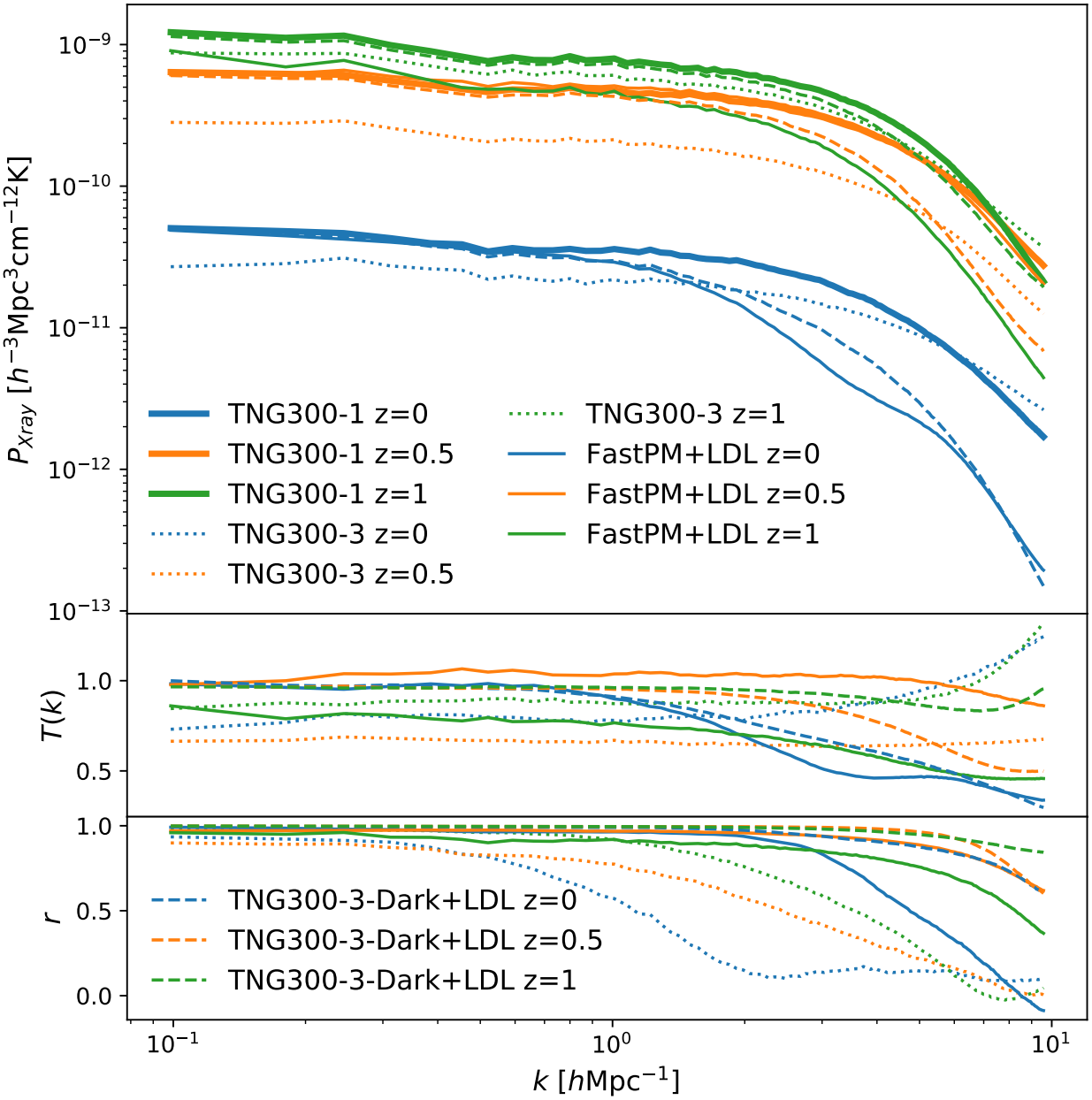



Figure 5.5: Comparison of the test set 3D power spectrum (top panel), transfer function (middle panel) and cross correlation coefficient (bottom panel) of the gas property $n_e^2 T^{0.5}$ (proportional to X-ray emissivity) between the LDL hybrid simulations, TNG300-3 and the target TNG300-1 hydrodynamical simulation.

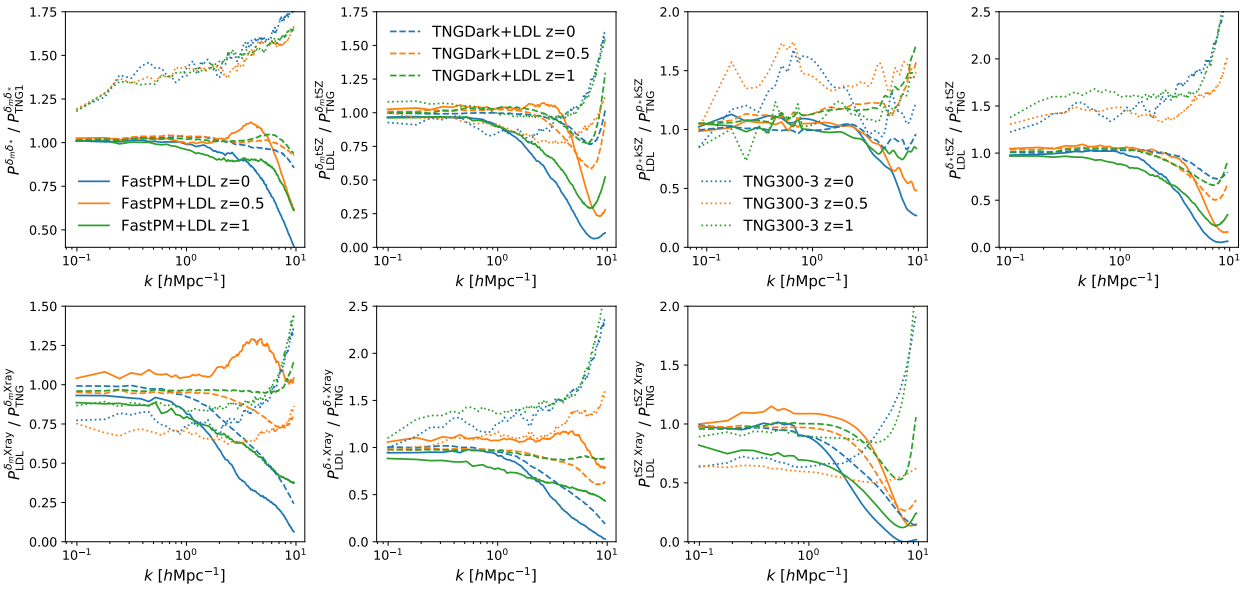



Figure 5.6: The ratio of the test set cross power spectrum of different observables between the LDL hybrid simulations and the target TNG300-1 hydrodynamical simulation. The first row show the cross power spectrum of matter and stellar mass (1st panel), matter and tSZ signal (2nd panel), 2D stellar momentum density and kSZ signal (3rd panel), and stellar mass and tSZ signal (4th panel). The second row shows the cross power spectrum of matter and X-ray (1st panel), stellar mass and X-ray (2nd panel), and tSZ and X-ray (3rd panel).

coefficient, which quantifies the agreement of phases of Fourier modes, is a more important statistics than the transfer function, because the transfer function can always be corrected to unity by multiplying the predicted fields with the reciprocal of the transfer function. This again suggests that the baryon maps of our models are closer to the ground truth than full hydrodynamical simulations at the same resolution.

## Cross Correlations between different tracers

Probes of the large-scale structure, such as weak lensing, galaxy survey and clusters, are strongly correlated because they are all determined by the same underlying matter distribution. There is additional information in the cross correlations between these probes which cannot be obtained by analyzing each observable independently. The cross correlation also has the advantage that the noise does not add to it. Our hybrid simulation is able to generate various observables simultaneously with a low computational cost, so it is potentially promising for cross correlation analysis. Here we investigate the predicted cross correlations between weak lensing convergence, mass weighted galaxies, tSZ and X-ray, as well as the cross correlation between the galaxy momentum and kSZ signal. We show the ratio of the predicted cross power spectrum and the TNG300-1 in Figure 5.6. Similar to the auto power spectrum analysis, our predicted cross power spectrum is in general

consistent with the target simulation on large scale (except for the cases involving X-ray predicted with FastPM, and we discuss possible reasons in the previous section), while TNG300-3 does not agree that well. On small scales FastPM-based hybrid simulation tends to underestimate the power, while TNG300-3 tends to overestimate the power. One can compare the second panel of Figure 5.6 (cross power spectrum between the matter and tSZ) with Figure 2 of [321], where GAN and VAE is used to predict the gas pressure from N-body simulations. We observe that the deviation of full-N-body-based hybrid simulation is comparable to the deviations of GAN and VAE. We note that for the standard deep learning architectures employed by GAN or VAE the number of parameters being fit is very large, in contrast to our approach.

## 5.4 Discussion

We propose a novel Lagrangian Deep Learning (LDL) model for learning the effective physical laws from the outputs of either simulations or real data. Specifically, in this paper we focus on learning the physics that controls baryon hydrodynamics in the cosmological simulations. We build hybrid simulations by combining N-body / quasi N-body gravity solver with LDL models. We show that both the FastPM-based and N-body-based hybrid simulations are able to generate maps of stellar mass, kSZ, tSZ and X-ray of various redshifts from the linear density field, and their computational costs are 7 and 4 orders of magnitudes lower than the target high resolution hydrodynamical simulation. We perform the auto power spectrum analysis and the cross correlation analysis among these fields, and we show that they generally outperform the hydrodynamical simulation at the same resolution.

The LDL model is motivated by the desire to provide an effective description of the underlying physics. Such a description must obey all the symmetries of the problem, and rotation and translation invariance are the two key symmetries, but other symmetries of the problem such as mass conservation may also appear. In this paper we argue that implementing these symmetries creates a generative model that is learning an effective description of the physical laws as opposed to learning the data distribution. This is because the symmetries are the only constraints on the generative model that must be implemented explicitly, everything else can be learned from the data. Here we propose that the learning of the generative model can be implemented by composing layers of displacements acting on the effective particles describing the physical properties of a system such as a fluid, moving the particles following the Lagrangian approach. The displacement of the particles can be understood as a result of the underlying physical processes, with particle transport a consequence of processes such as gas cooling and heating, feedback, turbulence etc. The output layer is a nonlinear transformation with thresholding on the particle density field, which models physics processes such as star formation.

Translational and rotational symmetry of the system put strong constraints on the model and therefore the Green's operator can be written as a function in Fourier Space that only depends on the amplitude of $k$. This allows us to use very few parameters to model the complex processes and produce maps of observables. Thus even though we want to describe systems of extremely high dimensionality ($10^8$ or more), the underlying effective physics description requires a handful

of parameters only.

The small number of free parameters also make the model stable and easy to train. An important advantage is that we can use the small number of parameters as an effective physics description of a complicated microphysics model, similar to the free parameters that arise from renormalization in the effective field theory descriptions of microphysics. This suggests that our LDL approach can replace other effective descriptions used to model the process of star formation. In cosmology such simplified models are often based on first identifying the dark matter halos in a dark matter simulation only, followed by some effective description of how to populate these halos with stars. Compared to such semi-analytical approaches which often rely on non-differentiable models, our approach is explicitly differentiable, such that we can use backpropagation to derive a gradient of the final observables with respect to the initial density field. This can be easily embedded into the forward modeling framework to reconstruct the initial conditions from the observations [287].

Our current implementation generally outperforms the full hydro simulation at the same resolution, but does not match perfectly the higher resolution hydro simulation. LDL deviates from the full simulation results mostly on small scales. This is expected, since the factor of 64 lower mass resolution means there is some information in the full simulation that cannot be recovered. Specifically, we use a low resolution mesh for calculating the displacements in the LDL model (cell size $0.328h^{-1}\mathrm{Mpc}$, see Table 5.1). The low resolution mesh limits the ability of LDL to model the small scale baryon distribution. Moreover, to ensure the correct large scale distribution, we apply a smoothing operator (Equation 5.9) to the fields before calculating the loss function, which downweights the small scale contribution to the loss function.

LDL trains on hydrodynamic simulations and is not meant to replace but to complement them: for example, it can interpolate a coarse grid and scale them to larger volumes and higher resolutions. In contrast, LDL has the potential to eliminate the need for the semi-analytic methods, which are the current standard paradigm in the large scale structure. These methods run N-body simulations first and then populate their dark matter halos using a semi-analytic prescription for the observable. LDL can not only achieve results that are on par with the full hydro at the same resolution, which is superior to these semi-analytic approaches, it also achieves this with of order 10 time steps, in addition to up to 6 LDL layers, in contrast to $10^3$ in an N-body simulations. We expect this will lead both to development of realistic simulations that cover the full volume of the cosmological LSS surveys, and to analysis of these LSS surveys with LDL effective parameters as the nuisance parameters describing the astrophysics of the galaxy formation.

## 5.5 Appendix A: Materials and Methods

### Dataset

IllustrisTNG is a suite of cosmological magneto-hydrodynamical simulations of galaxy formation and evolution [252, 228, 208, 302, 232]. It consists of three runs of different volumes and resolutions: TNG50, TNG100 and TNG300 with sidelengths of $35h^{-1}\mathrm{Mpc} \approx 50\mathrm{Mpc}$, $75h^{-1}\mathrm{Mpc} \approx 100\mathrm{Mpc}$ and $205h^{-1}\mathrm{Mpc} \approx 300\mathrm{Mpc}$, respectively. IllustrisTNG follows the evo-

lution of the dark matter, gas, stars and supermassive black holes, with a full set of physical models including star formation and evolution, supernova feedback with galactic wind, primordial and metal-line gas cooling, chemical enrichment, black hole formation, growth and multi-mode feedback. The IllustrisTNG series evolves over a redshift range $z = 127$ to the present $z = 0$ in a $\Lambda$CDM cosmology, with parameters $\Omega_m = 0.3089$, $\Omega_b = 0.0486$, $\Omega_\Lambda = 0.6911$, $H_0 = 67.74\mathrm{km\ s^{-1}\ Mpc^{-1}}$, $\sigma_8 = 0.8159$ and $n_s = 0.9667$.

In this paper we train our models against TNG300-1, the highest resolution of the TNG300 run. TNG300-1 evolves $2500^3$ dark matter particles and an initial number of $2500^3$ gas cells, with a comoving force resolution $\epsilon_{\mathrm{DM,stars}} = 1.0h^{-1}\mathrm{kpc}$, $\epsilon_{\mathrm{gas,min}} = 0.25h^{-1}\mathrm{kpc}$ and $\epsilon_{\mathrm{BH,max}} = 5.84h^{-1}\mathrm{kpc}$. The dark matter mass resolution is $4.0 \times 10^7 h^{-1} M_\odot$, and the target baryon mass resolution is $7.6 \times 10^6 h^{-1} M_\odot$ (see Table 5.1).

We also compare the model performance with TNG300-3, the hydro run with the same resolution as our hybrid simulations. The mass resolution and force resolution of TNG300-3 are 64 and 4 times lower than TNG300-1, respectively.

## Details of the Hybrid Simulation

The 10-step FastPM is run in a $205h^{-1}\mathrm{Mpc}$ periodic box, but with only $N = 625^3$ particles and force resolution $B = 2$. We generate the initial condition at redshift $z = 9$ using second order Lagrangian perturbation theory (2LPT), with the same random seed and linear power spectrum as Illustris-TNG. The linear density map is generated with a $N = 1250^3$ mesh to improve the accuracy on small scale [58]. The box is then evolved to redshift $0$ with 10 time steps that are linearly separated in scale factor $a$. Three snapshots are produced at redshift $z = 0$, $0.5$ and $1$, which are passed to LDL for generating maps of baryonic observables at these redshifts. Note that our mass, force and time resolutions are 64, 164 and 620,000 times lower than the target simulation TNG300-1, respectively.

Instead of running 10-step FastPM, we also tried using the particle data from the full N-body simulation TNG300-3-Dark. TNG300-3-Dark is the dark-matter-only run of the low resolution TNG300-3. It includes $N = 625^3$ dark matter particles (same as our FastPM setup), but the force and time resolution is significantly higher. A detailed comparison between FastPM, TNG300-3-Dark and TNG300-1 can be found in Table 5.1.

The details of the LDL model are described in the main text. We use a $N = 625^3$ mesh for calculating the displacement and generating the hydro maps. The architecture of the model is shown in Table 5.2. Specifically, for FastPM input, we firstly add a Lagrangian displacement layer and the output is matched to the density field of the full N-body simulation TNG300-3-Dark. This layer is intended to improve the small scale structure of FastPM and is shared by all hydro outputs (we do not add this layer for TNG300-3-Dark input). Then for different observables, we train different displacement layers and output layer: 1. For stellar mass, we add two displacement layer to mimic gas cooling and collapse, and one output layer to model star formation. 2. For kSZ signal, we use one displacement layer and one output layer to model the electron number density field. We assume that the velocities of gas trace dark matter, so the velocity field can be directly estimated from the dark matter particles: $v(x) = \frac{p(x)}{\rho(x)}$, where $p(x)$ is the momentum density field

Table 5.2: The LDL architecture for predicting different baryon observables

| | stellar mass | kSZ | | tSZ | | X-ray | |
|---|---|---|---|---|---|---|---|
| | | $n_e$ | $v_z$ | $n_e$ | $T$ | $n_e$ | $T$ |
| Displacement Layer (Eq. 5.1) | 2 | 1 | 0 | 1 | 2 | 2 | 2 |
| Output Layer (Eq. 5.5) | 1 | 1 | 0 | 1 | 1 | 1 | 1 |
| Total number of layers | 3 (4) | 2 (3) | | 5 (6) | | 6 (7) | |
| Total number of free parameters | 13 (18) | 8 (13) | | 21 (26) | | 26 (31) | |

For FastPM-based hybrid simulation, we add one more displacement layer to improve the small scale dark matter distribution. The corresponding $N_{\rm layer}$ and $N_{\rm parameter}$ are shown in parentheses.

and $\rho(x)$ is the matter density field. The kSZ map is obtained by multiplying the electron density field and the velocity field. 3. For tSZ signal map, we generate the electron number density field with one displacement layer and one output layer, and generate the gas temperature map with two displacement layers and one output layer. Then the two fields are multiplied to produce the tSZ signal. 4. The modeling of X-ray is similar to tSZ, except that now we use two displacement layer to model the electron density.

## Model Training and Loss Function

As described above, the output of the LDL model is a $N = 625^3$ mesh. We retain $77.7\%$ of the pixels for training, $13.8\%$ for validation and $8.5\%$ for test. Similar to [364], we split between training, validation, and test set following a "global" cut. The test set forms a sub-box of $90.2h^{-1}{\rm Mpc}$ per side, and the validation set is a $90.2 \times 114.8 \times 114.8\, h^{-1}{\rm Mpc}$ sub-box. The rest of the $205h^{-1}{\rm Mpc}$ box is used for training.

For stellar mass and the electron number density field in kSZ map, we define the loss function as:

$$\mathcal{L} = \sum_{i=1}^{N} \|\hat{\mathbf{O}}_{\mathbf{s}} F_{\rm LDL}(x_i) - \hat{\mathbf{O}}_{\mathbf{s}} F_{\rm TNG}(x_i)\|, \tag{5.8}$$

where $\|$ is $L_1$ norm, $i$ labels the mesh cell, $F_{\rm LDL}(x)$ is the generated map from LDL, $F_{\rm TNG}(x)$ is the true hydro map from IllustrisTNG, and $\hat{\mathbf{O}}_{\mathbf{s}}$ is a smoothing operator defined in Fourier space:

$$\hat{\mathbf{O}}_{\mathbf{s}} = 1 + (\frac{k}{1h{\rm Mpc}^{-1}})^{-n}. \tag{5.9}$$

Here $n$ is a hyperparameter that determines the relative weight between the large scale modes and the small scale modes. Without the $\hat{\mathbf{O}}_{\mathbf{s}}$ operator, the model focuses on the small scale distribution and results in a biased large scale power due to the small number of large scale modes relative to small scale modes. We apply $\hat{\mathbf{O}}_{\mathbf{s}}$ operator to put more weight on the large scale distribution. For most of the baryon maps we fix $n = 1$, except for the X-ray map we optimize the hyperparameter $n$.

For the tSZ map, we use a different loss function to improve the performance. We firstly train the electron density map with the following loss function:

$$\mathcal{L}_{n_e}^{\rm tSZ} = \sum_{i=1}^{N} \|\hat{\mathbf{O}}_{\mathbf{s}}[n_{e_{\rm LDL}}(x_i)T_{\rm TNG}(x_i)] - \hat{\mathbf{O}}_{\mathbf{s}}[n_{e_{\rm TNG}}(x_i)T_{\rm TNG}(x_i)]\|, \tag{5.10}$$

where $n_{e_{\rm LDL}}(x)$ is the learned electron number density map, $n_{e_{\rm TNG}}(x)$ is the true electron number density map, and $T_{\rm TNG}$ is the true temperature map. This means we multiply the electron number density field with the temperature field before calculating the loss function. This procedure puts more weight on the large clusters and improves the quality of the generated tSZ maps. Note that this electron density field is different from the electron density field for predicting the kSZ signal. Similarly, after we obtain the learned electron number density field $n_{e_{\rm LDL}}(x)$, we train the temperature map with the following the loss function:

$$\mathcal{L}_{T}^{\rm tSZ} = \sum_{i=1}^{N} \|\hat{\mathbf{O}}_{\mathbf{s}}[n_{e_{\rm LDL}}(x_i)T_{\rm LDL}(x_i)] - \hat{\mathbf{O}}_{\mathbf{s}}[n_{e_{\rm TNG}}(x_i)T_{\rm TNG}(x_i)]\|. \tag{5.11}$$

Here $n_{e_{\rm LDL}}(x)$ is the electron density field we just learned and is fixed, and $T_{\rm LDL}(x)$ is the target temperature field that we are trying to optimize.

For the X-ray map, similar to the tSZ signal, we train the electron density and gas temperature maps successively with the following loss functions:

$$\mathcal{L}_{n_e}^{\rm X} = \sum_{i=1}^{N} \|\hat{\mathbf{O}}_{\mathbf{s}}[n_{e_{\rm LDL}}^2(x_i)T_{\rm TNG}^{0.5}(x_i)] - \hat{\mathbf{O}}_{\mathbf{s}}[n_{e_{\rm TNG}}^2(x_i)T_{\rm TNG}^{0.5}(x_i)]\|, \tag{5.12}$$

$$\mathcal{L}_{T}^{\rm X} = \sum_{i=1}^{N} \|\hat{\mathbf{O}}_{\mathbf{s}}[n_{e_{\rm LDL}}^2(x_i)T_{\rm LDL}^{0.5}(x_i)] - \hat{\mathbf{O}}_{\mathbf{s}}[n_{e_{\rm TNG}}^2(x_i)T_{\rm TNG}^{0.5}(x_i)]\|. \tag{5.13}$$

Again, the electron number density field and gas temperature field for X-ray are different from the fields used for generating kSZ and tSZ.

Because the number of free parameters is relatively small, in this work we use the L-BFGS-B algorithm [36] for optimizing the model parameters. The training time varies with the target. A typical training (e.g. stellar mass) takes a couple of hours on 4 NERSC Cori nodes, which corresponds to a few hundred CPU hours.

# Acknowledgements

We thank Dylan Nelson and the IllustrisTNG team for kindly providing the linear power spectrum, random seed and the numerical parameters of the IllustrisTNG simulations. We thank Yu Feng for helpful discussions. The majority of the computation were performed on NERSC computing facilities Cori, billed under the cosmosim and m3058 repository. National Energy Research

Scientific Computing Center (NERSC) is a U.S. Department of Energy Office of Science User Facility operated under Contract No. DE-AC02-05CH11231. The power spectrum analysis in this work is performed using the open-source toolkit nbodykit [118].

# Chapter 6

# A field-level emulator for modeling baryonic effects across hydrodynamic simulations

The contents of this chapter were originally published in [292],

A field-level emulator for modeling baryonic effects across hydrodynamic simulations
*Sharma D., Dai B., Villaescusa-Navarro F., Seljak U.*(arXiv:2401.15891) (submitted to MNRAS)

We develop a new and simple method to model baryonic effects at the field level relevant for weak lensing analyses. We analyze thousands of state-of-the-art hydrodynamic simulations from the CAMELS project, each with different cosmology and strength of feedback, and we find that the cross-correlation coefficient between full hydrodynamic and N-body simulations is very close to 1 down to $k \sim 10\ h\mathrm{Mpc}^{-1}$. This suggests that modeling baryonic effects at the field level down to these scales only requires N-body simulations plus a correction to the mode's amplitude given by: $\sqrt{P_{\rm hydro}(k)/P_{\rm nbody}(k)}$. In this paper, we build an emulator for this quantity, using Gaussian processes, that is flexible enough to reproduce results from thousands of hydrodynamic simulations that have different cosmologies, astrophysics, subgrid physics, volumes, resolutions, and at different redshifts. Our emulator is accurate at the percent level and exhibits a range of validation superior to previous studies. This method and our emulator enable field-level simulation-based inference analyses and accounting for baryonic effects in weak lensing analyses.

## 6.1 Introduction

Weak gravitational lensing is a powerful tool for measuring the clustering of matter in our universe, and thus obtaining information about matter content and initial conditions of our universe [159] through various summary statistics. However, deriving precise cosmological constraints from weak lensing observations necessitates highly accurate theoretical models that account for baryonic physics, which redistributes matter on small scales via processes such as Active Galactic Nuclei

(AGN) feedback. These processes remain poorly understood and inadequately constrained by current observations, leading to challenges in formulating a predictive theory.

While causality ensures that baryonic effects are negligible on large scales [182], baryonic effects gain significance on smaller scales, affecting structure formation through hydrodynamic processes. These processes, including AGN and stellar feedback, can heat gas and inject large amounts of energy into galaxies and the surrounding halo. Specifically, AGN feedback can eject gas to very large distances, which can further modify the dark matter distribution through gravitational interactions.

Studies have shown that probing the small scales contains a wealth of information, leading to stronger parameter constraints [196], while ignoring small scale information leads to significant deterioration of these constraints [173, 174, 131]. This small-scale information is heavily influenced by baryonic effects, and the large uncertainty associated with these effects makes them one of the primary sources of systematic error in weak lensing analyses. Hence, accurate modeling of baryonic effects is crucial when probing the information-rich small scales for an unbiased cosmological analysis.

At present, numerical simulations remain the sole comprehensive method for precise simulation of baryonic effects and the deeply non-linear evolution of cosmic structures. Hydrodynamic simulations provide a detailed and accurate representation of the behavior of baryonic matter by modeling complex physical processes such as gas dynamics, star formation, and feedback mechanisms [316, 67, 150]. However, these simulations are not ab initio parameter free, but instead must parametrize the lack of physics understanding via free parameters that can be varied. Furthermore, hydrodynamic simulations require substantial computational resources as compared to dark matter only N-body simulations.

The development of emulators has emerged as a powerful technique to overcome this computational challenge, enabling rapid and accurate predictions of physical properties without the need for running costly simulations. These emulators interpolate simulation results and have been shown to be remarkably accurate. Various emulators have been developed for cosmology, catering to various observables, encompassing the matter power spectrum [126, 84, 354, 10, 170], mass function [217, 27], and the galaxy correlation function and Lyman-$\alpha$ Forest [363, 23].

While most previous work has focused on modeling the baryonic effects on the matter power spectrum [138, 219, 12, 283, 282, 104], there is an increasing need for developing fast baryon models at the field level for analysis beyond two-point statistics. For example, simulation-based inference methods [51] show great promise in extracting rich non-Gaussian information either through high-order statistics [e.g., 114], or directly from the fields [e.g., 54, 56, 333, 335]. These approaches rely on fast and accurate cosmological predictions from numerical simulations. Previous field-level baryon models, such as Baryon Correction Model [281] and Enthalpy Gradient Descent [53], move the dark matter particles from N-body simulations to mimic the baryonic effects. While they have been shown to accurately predict the power spectrum from hydrodynamical simulations [283], they can be computationally expensive when the particle resolution is high.

By analyzing a diverse range of baryonic feedback hydrodynamics simulations across multiple redshifts, we will show that adding baryons to N-body simulations can be achieved using a field-level transfer function to augment N-body fields with a Fourier mode amplitude, k, dependent

transfer function correction. We develop a transfer function emulator using Gaussian process for modeling the baryonic effects in terms of $P_{\rm hydro}(k)/P_{\rm nbody}(k)$, where $P_{\rm hydro}(k)$ is the total matter power spectrum, and $P_{\rm nbody}(k)$ is the dark matter power spectrum.

In this paper, we develop the emulator and show that it is accurate at a percent level over our whole parameter space, which covers scales $0.01 \leq k \leq 10$ h/Mpc and redshifts $0 \leq z \leq 1.5$. We validate the performance of our emulator against thousands of hydrodynamical simulations and their respective gravity-only counterparts. In particular, we make use of CAMELS-Astrid [237], CAMELS-IllustrisTNG, CAMELS-SIMBA[337, 338], BAHAMAS [215, 216], Horizon AGN [78], Owls [279, 326], and Eagle [278, 50, 212, 127] simulations. We also compare against commonly utilized emulators like BACCO [12], HMcode [219], and BCemu [104] on all simulations. Finally, we show the improvement of the field-level baryon model against hydrodynamic fields at varying redshifts. This emulator is fast as it only requires a single FFT and its inverse, which enables large-volume N-body simulations for generating realistic weak lensing mock data for cosmological analysis at the field level.

This paper is organized as follows: in section 6.2, we describe the suite of simulations that are used for training and testing our emulator. In section 6.3 we explain how our emulator can be used to emulate baryonic effects at the field level. In section 6.4 we describe the methods and construction of the Gaussian process emulator. In section 6.5 we test the emulator's robustness on multiple hydrodynamic test simulations, compare it with currently available emulators, and show field-level improvements using our emulator. We summarize and conclude in section 6.6.

## 6.2 Simulations

In this section, we describe the simulations that we employ throughout this paper. Our main suites of simulations are part of the Cosmology and Astrophysics with MachinE Learning Simulations (CAMELS) [337, 338, 237]. CAMELS is a suite of 10,421 cosmological simulations each with a comoving volume of $(25\ h^{-1}{\rm Mpc})^3$ evolved from $z = 127$ to $z = 0$ with $256^3$ dark matter particles and $256^3$ gas particles in the initial conditions. These contain 5,097 N-body simulations and 5,324 hydrodynamic simulations. Notably, each hydrodynamic simulation in CAMELS pairs with an N-body counterpart, sharing identical cosmological parameters and initial random seeds.

Simulations in CAMELS are categorized into various suites (Astrid, IllustrisTNG, and SIMBA) and sets based on the employed code for running the simulations and the arrangement of cosmological and astrophysical parameters ($\Omega_m, \sigma_8, A_{\rm SN1}, A_{\rm AGN}, A_{\rm SN2}, A_{\rm AGN2}$), as well as the initial random seeds. The Astrid suite comprises 1,092 hydrodynamic simulations executed using the MP-Gadget simulation code [91], employing analogous subgrid physics as the original Astrid simulations [236, 24]. Additionally, the IllustrisTNG suite [based on 343, 320] and the SIMBA suite [based on 60], with 1,092 hydrodynamic simulations each, are executed using the AREPO code [299, 351] and the GIZMO code [135], respectively.

Each simulation is characterized by its cosmology (given by $\Omega_m$ and $\sigma_8$) and its astrophysical feedback (given by $A_{\rm SN1}, A_{\rm AGN}, A_{\rm SN2}, A_{\rm AGN2}$). In particular, throughout CAMELS' suites, the

astrophysical parameters represent the value of subgrid physics parameters that influence stellar and Active Galactic Nuclei (AGN) feedback mechanisms.

We made use of the Latin Hypercube, LH, set within each suite of CAMELS, which contains 1,000 simulations whose cosmological and astrophysical parameters are arranged in a latin-hypercube within a very broad range[1]

$$\Omega_m \in [0.1, 0.5] \tag{6.1}$$
$$\sigma_8 \in [0.6, 1.0] \tag{6.2}$$
$$A_{\rm SN1}, A_{\rm AGN1} \in [0.25, 4.0] \tag{6.3}$$
$$A_{\rm SN2}, A_{\rm AGN2} \in [0.5, 2.0] \tag{6.4}$$

and every simulation has a different value of the initial random seed. Within each LH set, CAMELS provides 1,000 simulations for each redshift that span a wide range of cosmologies and baryonic feedbacks, perfect for our purposes of capturing the underlying physics.

Importantly, each suite has been run with a different code and therefore the subgrid physics model is completely different. So, notably, while the range of variation of the above parameters remains consistent across all CAMELS suites, the precise definitions and overall impact of these astrophysical parameters vary significantly across suites. These simulations are designed to train and set machine learning algorithms given the way their cosmological, astrophysical, and initial random seed parameters are set. We utilize this set for each suite throughout this paper to train and test our methodology on simulations with significantly different cosmologies and astrophysics.

Figure 6.1 shows the baryonic effect in the matter power spectrum, $P_{\rm hydro}(k)/P_{\rm nbody}(k)$, across different redshift values in all CAMELS suites utilized in this study. From the figure, it is clear that baryonic feedback can have very diverse and strong effects on the matter power spectrum, especially on small scales. SIMBA, with its aggressive AGN feedback, produces the most prominent suppression of the matter power on large scales. On the other hand, IllustrisTNG exhibits a more moderate impact on the matter power spectrum as a consequence of having milder AGN feedback, and Astrid spans the widest range of baryonic feedback, encompassing effects seen in both SIMBA and IllustrisTNG.

To cover the broadest range of baryonic feedback, based on Figure 6.1, we used simulations from the Astrid suite at $z = 0.0$ to train our emulator. For selecting the training simulations, we performed a random sampling of 800 simulations (out of the 1000 available) from the Astrid suite within CAMELS at $z = 0.0$.

Post-training, we test the emulator using the remaining 200 simulations from Astrid at $z = 0.0$ alongside all other available hydrodynamic simulation suites in CAMELS. Figure 6.2 illustrates the matter power spectrum ratio in the Astrid $z = 0.0$ test data. Additionally, the IllustrisTNG suite and the SIMBA suite are part of the test dataset.

Previous studies [126, 295, 265, 126] have shown that both high physical resolution and large box sizes are required to guarantee convergence of the power spectrum. [281] showed that deviations of the power spectrum ratio using small boxes, like in CAMELS, are at the $5\%$ level. However,

[1] We note that in the case of Astrid, the parameter $A_{\rm AGN2}$ varies between 0.25 and 4.

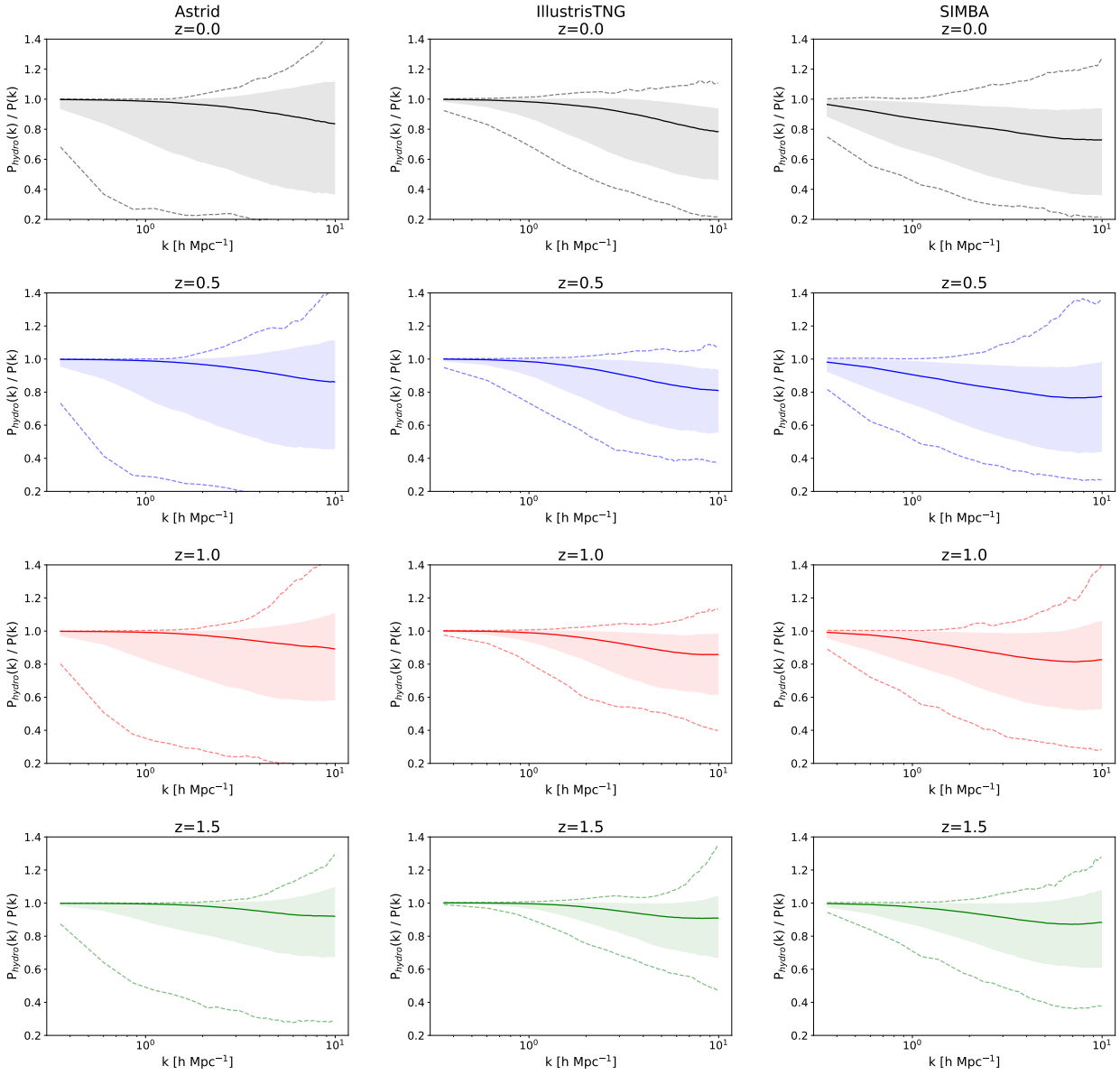



Figure 6.1: The matter power spectra ratio observed across suites within CAMELS at various redshifts. Each simulation suite's median value is represented by the solid line, while the dashed lines denote the extreme values, providing an overview of the suite's variability. The shaded region indicates 90 percentiles, reflecting the statistical distribution. During the training phase, our emulator exclusively utilizes 800 Astrid simulations at $z = 0.0$. Post-training, we test the emulator on all other CAMELS simulations shown here in addition to the remaining 200 Astrid $z = 0.0$ simulations in figure 6.2 and other simulations outside CAMELS in figure 6.6. Testing the emulator for such varied simulations serves to evaluate the emulator's reliability and generalizability across a wide spectrum of redshifts and simulations.

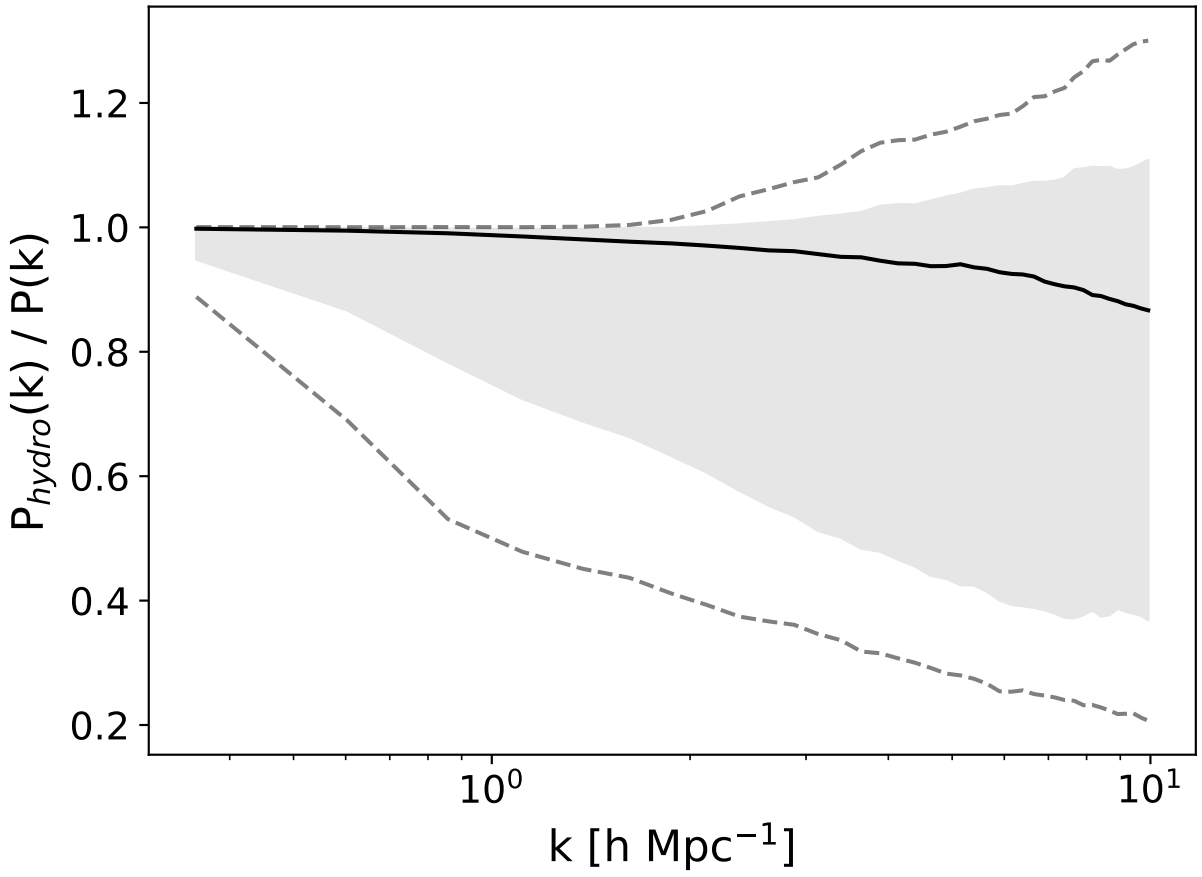



Figure 6.2: The ratio of matter power spectra in the 200 Astrid test simulations for $z = 0.0$. The solid line denotes the median value across the simulations, while the dashed lines illustrate the extreme values. The shaded region delineates the 90th percentile range. During emulator training, 800 simulations were randomly sampled from the Astrid (at $z = 0.0$) suite within CAMELS, leaving these 200 remaining simulations for post-training Astrid $z = 0.0$ test analysis.

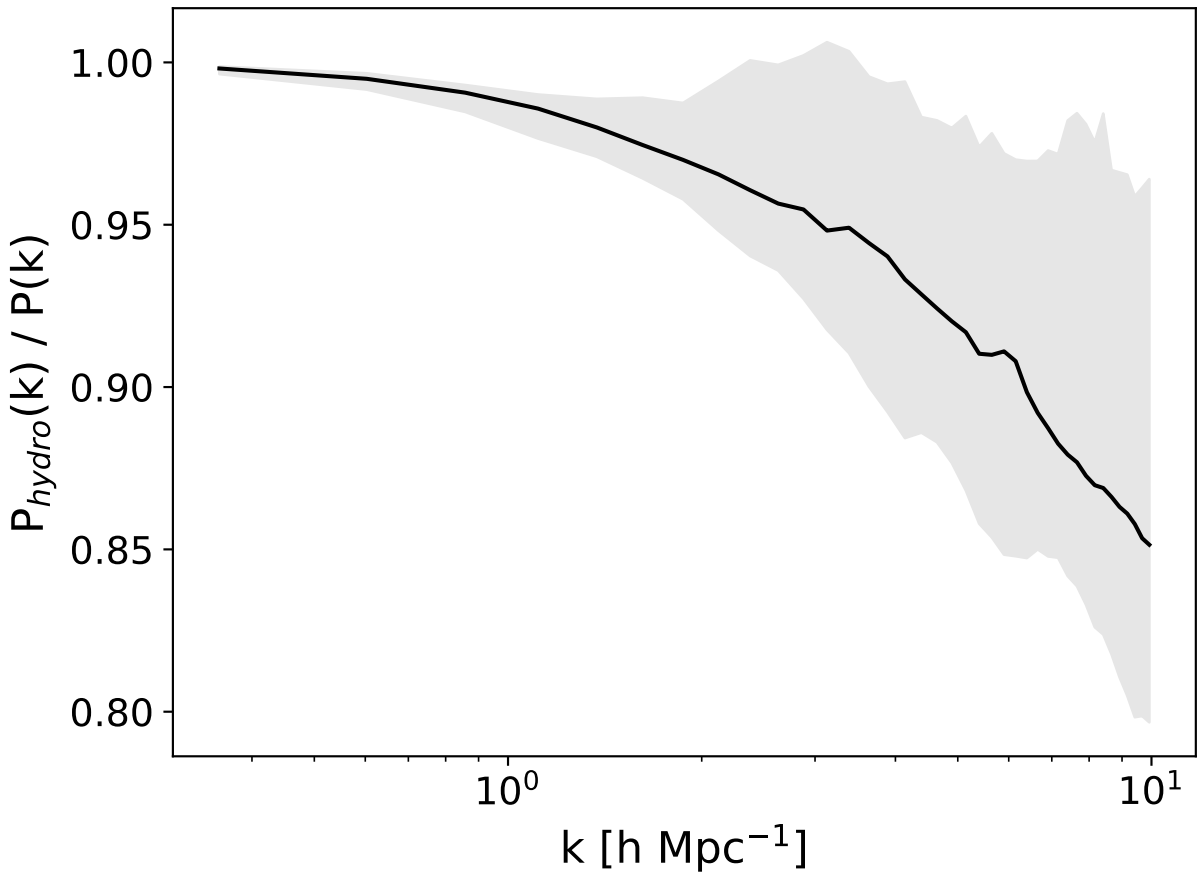



Figure 6.3: Ratios of the matter power spectra in the Astrid $z = 0.0$ CV set. The solid line shows the median value, while the shaded region shows the entire range of the ratios. All the simulations share the value of the cosmological and astrophysical parameters; they only differ in the value of their initial conditions random seed and hence can be used to study the effect of cosmic variance.

these could also be due to cosmic variance which affects the small CAMELS-like box volumes. To study the effects of cosmic variance on the matter power spectrum ratio, we show the ratios for simulations with the same cosmology and astrophysics from CAMELS-Astrid's CV set in figure 6.3. From this, we can see that the matter power spectrum ratio is affected by cosmic variance up to $\sim 10\%$ on small scales, suggesting that the deviations for small boxes are due to cosmic variance.

In addition to the diverse array of simulations in CAMELS, we extended the validation of our emulator by testing it against simulations outside the CAMELS database. These external simulations, including BAHAMAS [215, 216], Horizon AGN [78], Owls [279, 326], and Eagle [278, 50, 212, 127], serve as crucial benchmarks to assess the robustness and generalizability of the emulator's predictions beyond CAMELS. These simulations encompass various diverse physical

processes including AGN feedback, supernovae feedback, mass loss from Asymptotic Giant Branch stars, radiative cooling, stellar winds, and stellar initial mass function, among others. Additionally, these simulations differ from those in CAMELS in volume, resolution, and subgrid physics code [327], enabling us to test the robustness of our emulator to these effects. The solid lines in the left panel of Figure 6.6 illustrate the matter power spectrum ratio derived from these external simulations, allowing us to assess how the emulator performs across varied simulations beyond the scope of the CAMELS.

## 6.3 Baryonic effects at the field-level

In this section, we present our methodology to model baryonic effects at the field level for the total matter density field. We start by computing the cross-correlation coefficient between the matter field in full hydrodynamic simulations and their N-body counterparts. The cross-correlation coefficient, $r(k)$, is defined as:

$$r(k) = \frac{P_{\rm cross}(k)}{\sqrt{P_{\rm nbody}(k)P_{\rm hydro}(k)}} \tag{6.5}$$

Here, $P_{\rm cross}(k)$, $P_{\rm nbody}(k)$, and $P_{\rm hydro}(k)$ represent the cross-power spectrum, the N-body power spectrum, and the power spectrum of the hydrodynamic fields respectively. This coefficient's range spans from $-1$ to 1, where values closer to 1 signify a strong positive linear relationship, $-1$ indicates a strong negative linear relationship and 0 implies no linear relationship between the datasets.

Figure 6.4 shows the cross-correlation coefficients derived from all CAMELS suites at different redshift values pertinent to this study. These coefficients serve as indicators of the correlation strength between N-body and hydrodynamic fields within the simulations.

We can see that the calculated cross-correlation coefficients are very close to 1, down to $k \sim 10$ h/Mpc, for all the simulations, with most deviations being within $5 - 10\%$. On the other hand, we see in Figure 6.1 that baryonic effects can cause deviations of up to $\sim 50\%$ on the matter power spectrum, with the effects getting more dominant at smaller scales. This suggests that the baryonic effects predominantly impact the amplitude of the Fourier modes (given by the power spectrum) rather than their phases (given by cross-correlation coefficients).

Since the amplitude and phase of the Fourier modes completely describe the fields, with baryonic effects mainly changing the amplitudes, aligning the power spectra of N-body fields with their hydrodynamic counterparts would also effectively align them at the field level, facilitating a cost-effective field-level analysis using just N-body simulations.

We can achieve this power spectra alignment by applying a transfer function to the N-body fields [31, 289]. Transfer functions operate by performing specific modifications to the field data. In Fourier space, each mode of a field is represented by its amplitude, typically denoted by $|\mathbf{k}|$, and its phase. Transfer functions act on $\mathbf{k}$ to alter their amplitudes according to certain criteria [242, 73].

The field-level transformation using a transfer function, $T(k)$, is mathematically defined as:

$$F'(\mathbf{k}) = T(k) \cdot F(\mathbf{k}) \tag{6.6}$$

Here, $F(\mathbf{k})$ symbolizes the original field in Fourier space, and $F'(\mathbf{k})$ represents the transformed field in Fourier space after the element-wise application of the transfer function $T(k)$ to the original field $F(\mathbf{k})$. This transformation enables the adjustment of simulated fields to match desired characteristics or observational data, enhancing the accuracy or realism of the simulation results.

In our context of incorporating baryonic effects in N-body simulations, we apply a transfer function to the simulated field to adjust its power spectrum. Defining baryonic suppression as:

$$S(k) \coloneqq P_{\rm hydro}(k)/P_{\rm nbody}(k) \tag{6.7}$$

Our field transformation on N-body fields is then:

$$\delta'_{\rm nbody}(\mathbf{k}) = \sqrt{S(k)} \cdot \delta_{\rm nbody}(\mathbf{k}) \tag{6.8}$$

This could increase or suppress the power of certain scales and correct discrepancies arising from missing baryonic physics effects in N-body simulations, aligning the power spectra of N-body fields with the full hydrodynamic fields.

## 6.4 Gaussian Process Emulator

To fulfill the promise of field-level modeling of baryonic effects, we need to characterize $P_{\rm hydro}(k)/P_{\rm nbody}(k)$ for our transfer function. In this section, we describe the main numerical methods we use to create our emulator of $P_{\rm hydro}(k)/P_{\rm nbody}(k)$ using Gaussian Processes.

Gaussian Processes (GPs) [266] are a versatile tool within machine learning and statistics, renowned for their efficacy in regression, interpolation, and uncertainty quantification. They provide a flexible framework for modeling functions along with their associated uncertainties [e.g. 23, 273, 272, 243]. Moreover, a Gaussian process emulator is computationally efficient, enabling its use within standard inference methodologies such as Markov Chain Monte Carlo (MCMC) for evaluations.

For the length scales we want to model, the CAMELS simulations have 39 linearly-spaced $k$ bins spanning the range $0.36 < k < 9.93$ h/Mpc. Hence, to model baryonic effects at small scales down to $k \sim 10$, our emulator delineates these 39 linearly-spaced $k$ bins. We treat the baryonic effects in each $k$ bin as an individual Gaussian process, enabling independent training for each bin, with every simulation serving as a training point for these $k$ bins.

At a specific point $X = [\Omega_m, \sigma_8, A_{\rm SN1}, A_{\rm AGN}, A_{\rm SN2}, A_{\rm AGN2}]$ in the parameter space, a Gaussian process models the target function — $S(\mathbf{k}|X) \coloneqq P_{\rm hydro}(\mathbf{k}|X)/P_{\rm nbody}(\mathbf{k}|X)$ in our case — as an assembly of random variables that form a joint Gaussian distribution. This model is defined via $S(\mathbf{k}|X) \sim \mathcal{N}(0, K(X, X_i))$, where $X_i$ signifies the parameter values at the training simulations, and $K(X, X')$ represents a covariance kernel.

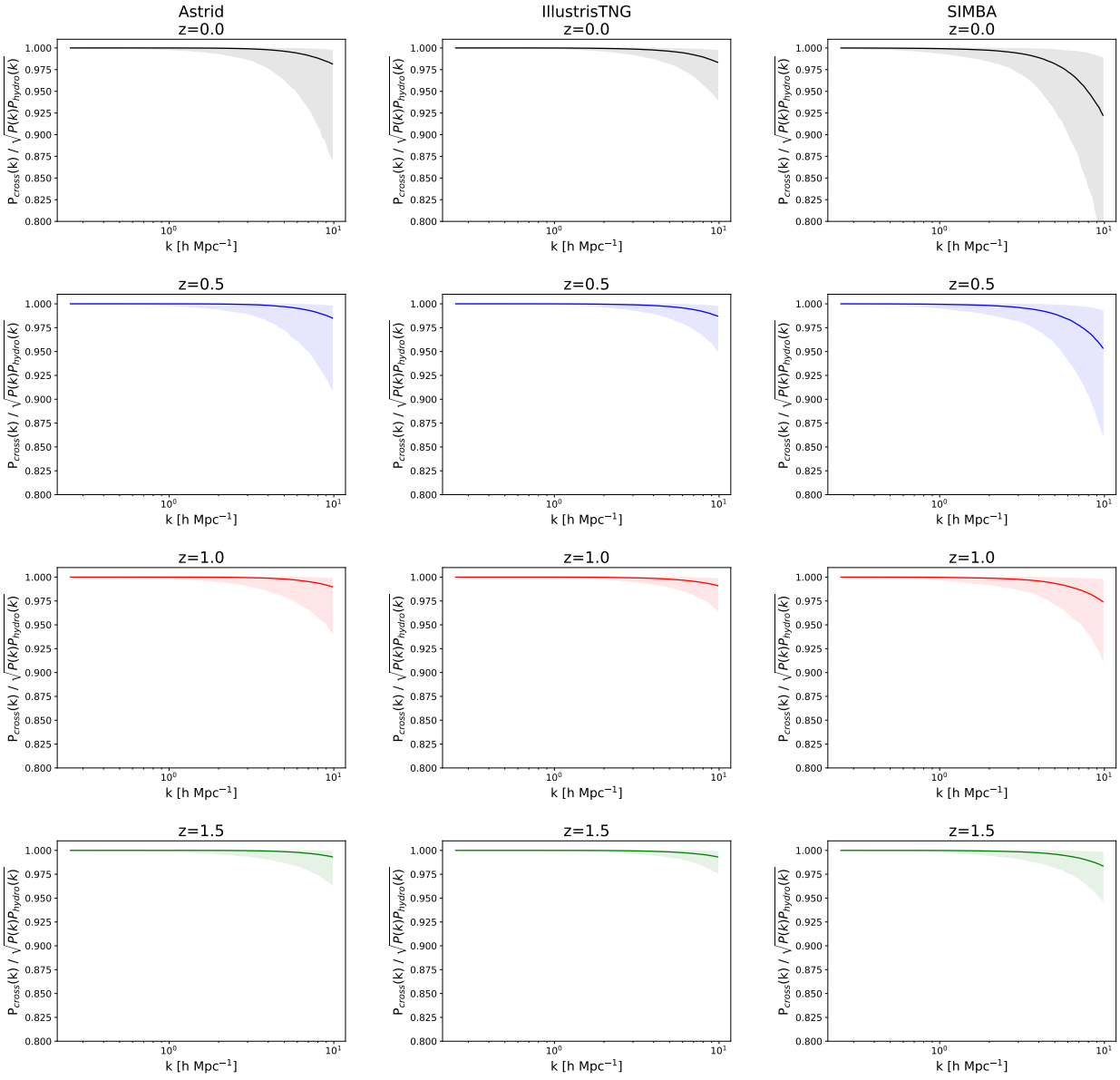



Figure 6.4: The cross-correlation coefficients between N-body and hydrodynamic fields across all suites within CAMELS at different redshifts. The solid line signifies the median value, while the shaded region delineates the 90th percentile range, reflecting the distribution and variability of these coefficients. Remarkably, the computed cross-correlation coefficients for all simulations are very close to 1, denoting a strong relationship between N-body and hydrodynamic fields. This suggests that the impact of baryonic effects predominantly alters the amplitude (power) of the Fourier modes while exerting minimal influence on their phase (cross-correlation coefficient). Consequently, we posit that field-level enhancements in N-body fields can be achieved by refining their power spectra to closely align with the corresponding hydrodynamic fields. Targeting the power spectra improvement holds promise for effectively reconciling the discrepancies between N-body and hydrodynamic simulations, facilitating more accurate and cost-effective field-level analyses across simulations.

The choice of the kernel function $K(X, X_i)$ plays a pivotal role in characterizing the correlation or similarity between data points $X$ and $X_i$. This function serves as a prior, encapsulating the expected behavior of the underlying function — baryonic effects in our case — to be modeled. For our emulator, we adopt a Matérn 5/2 kernel, a generalized form of the radial basis function (RBF), defined by:

$$K(X, X') = \sigma_0^2 \left( \sum_{i=1}^{6} \left( 1 + \frac{\sqrt{5}r}{\ell_i} + \frac{5r^2}{3\ell_i^2} \right) \exp\left( -\frac{\sqrt{5}r}{\ell_i} \right) \right) \tag{6.9}$$

Here, $r = ||X - X'||_2$ represents the L2 distance between two data points, $\sigma_0^2$ denotes the variance parameter, and $\ell_i$ signifies the length scale of each input dimension, influencing the smoothness and range of correlations between data points. Our choice of this covariance kernel is motivated by the need for flexibility, achieved through the squared exponential kernel, the efficacy of linear interpolation, and allowing for noise in the training data.

At a test point $X_*$, the joint distribution of the test data $S(\mathbf{k}|X_*)$ and training data $S(\mathbf{k}|X_i)$ can be expressed as:

$$\begin{bmatrix} S(\mathbf{k}|X_i) \\ S(\mathbf{k}|X_*) \end{bmatrix} \sim \mathcal{N}\left( 0, \begin{bmatrix} K(X_i, X_i) + \sigma_n^2 I & K(X_i, X_*) \\ K(X_*, X_i) & K(X_*, X_*) \end{bmatrix} \right) \tag{6.10}$$

Here, $\sigma_n^2$ serves as a hyperparameter signifying Gaussian noise within the training data.

Consequently, our Gaussian process involves a total of 8 hyperparameters: 6 correlation lengths, $\sigma_0^2$, and $\sigma_n^2$. The hyperparameters are optimized by maximizing the marginal log-likelihood of the training data [266].

The posterior predictive distribution over test data is obtained through:

$$S(\mathbf{k}|X_*) \sim \mathcal{N}(\mu, \Sigma) \tag{6.11}$$

$$\mu = K(X_*, X_i) \left( K(X_i, X_i) + \sigma_n^2 I \right)^{-1} S(\mathbf{k}|X_i), \tag{6.12}$$

$$\Sigma = K(X_*, X_*) - K(X_*, X_i) \left( K(X_i, X_i) + \sigma_n^2 I \right)^{-1} K(X_i, X_*) \tag{6.13}$$

The mean and variance derived from the posterior predictive distribution, using the training information at $X_i$, serve as estimators for the value and interpolation uncertainty associated with $S(X_*)$.

We implement our emulator using tinygp [96], a Python library for GP Regression (GPR) built on top of the JAX library for numerical computing [35].

The GP model offers a broad prior across function space, enabling the modeling of the diverse baryonic effects we see in figure 6.1 without imposing strong prior constraints on its parameter dependencies. Since it is stochastic, this model provides predictions for the baryonic suppression beyond the training points, accompanied by associated uncertainties that can be integrated into

our statistical model. This emulation methodology provides a robust approach for modeling and predicting the baryonic effects in simulations, enabling efficient and accurate interpolation, and quantification of uncertainties.

## 6.5 Results

In this section, we use the trained Gaussian Process emulator to generate predictions for $P_{\rm hydro}(k)/P_{\rm nbody}(k)$ as a function of four astrophysical parameters - $A_{\rm SN1}, A_{\rm AGN}, A_{\rm SN2}, A_{\rm AGN2}$ within their respective ranges. We emphasize that our emulator was trained on Astrid simulations at $z = 0$, and therefore, the meaning of these astrophysical parameters is, in principle, associated with the Astrid subgrid physics model. However, to make our emulator generic and robust, from now on, we will consider these four astrophysical parameters as nuisance parameters that one needs to tune to reproduce the result of one particular hydrodynamic simulation.

We employ the differential evolution global optimizer from the SciPy library [340] to obtain the best-fit value of these nuisance parameters. This optimization technique is adept at exploring the parameter space to seek optimal solutions, especially in scenarios with complex, multi-dimensional parameter spaces. The differential evolution [305, 306, 260] method operates stochastically, offering a non-gradient approach to locating the minimum and can search through large volumes in parameter space.

We now show the accuracy of our emulator for simulations within and outside CAMELS, showing its precision to changes in simulation cosmology, feedback, subgrid physics, resolution, volume, and redshift. On top of that, we compare the accuracy of our emulator against other emulators in the literature. Finally, we demonstrate the emulator's efficacy in creating field-level improvements when applied to the N-body fields of simulations within CAMELS, validating its potential for advancing field-level weak-lensing analyses using cosmological simulations.

### Emulator accuracy

We start by quantifying the accuracy of our emulator across hydrodynamic simulations.

- **CAMELS simulations.** Figure 6.5 shows the error achieved by our emulator for simulations of three different suites of CAMELS (IllustrisTNG, Astrid, and SIMBA) at four different redshifts. The solid lines represent the average percent error across simulations, while the shaded regions denote the 90th percentile range. These results correspond to all the baryonic effects illustrated in figure 6.1. Firstly, we can see that the emulator achieves a high accuracy down to $k \sim 10$ h/Mpc with deviations remaining typically less than 5%. We emphasize that our emulator is robust to changes in redshifts and baryonic effects across CAMELS.

  The performance of the emulator is similar across redshifts for the Astrid and IllustrisTNG simulations at all scales, with higher accuracy at large scales and somewhat lower precision at smaller scales. However, at $z = 0.0$ and $0.5$, the prediction error for SIMBA can be as high $\sim 5\%$ on large scales. This is likely due to the aggressive AGN feedback in SIMBA, which

produces the most prominent suppression of the matter power on large scales [101] as seen even in the comparison plots in Figure 6.1. Nonetheless, the prediction error is still within $\sim 5\%$ and is comparable to the other two suites at $z = 1.0, 1.5$.

- **Non-CAMELS simulations.** While the above test shows the robustness of our emulator to changes in cosmology, astrophysics, and subgrid physics, we note that all CAMELS simulations share the same volume and resolution. In order to quantify how well our emulator behaves to changes in volume, resolution, and other subgrid physics models, we quantify how well it is able to reproduce the results of the BAHAMAS, Horizon AGN, Owls, and Eagle simulations. We show the results in Figure 6.6.

  The left panel shows the correction to the matter power spectrum in these simulations; solid lines represent the simulation results, while the corresponding dashed line depicts our emulator's predictions. The emulator closely mirrors the inherent behavior of baryonic effects across these varied simulations. In the right panel, the prediction errors for each simulation are displayed. Consistently, the emulator maintains accuracy at the percent level up to $k \sim 10$ h/Mpc, adeptly capturing the intricacies of baryonic effects across diverse scenarios. All the above tests clearly illustrate the versatility and robustness of our emulator, which is capable of reproducing the ratio $P_{\rm hydro}(k)/P_{\rm nbody}(k)$ for thousands of simulations with different cosmologies, astrophysics, subgrid physics, volumes, resolutions, and redshifts.

## Comparison against other emulators

In recent years, different groups have created emulators to model baryonic effects for 2-point statistics. Given the findings of this work, we can also use those emulators to create field-level baryonic effects corrections. In this subsection, we conduct comparative evaluations against widely used emulators such as BACCO, HMcode, and BCemu, both within and beyond the CAMELS simulations. Through the following comparisons, we show that, overall, our emulator offers greater flexibility and robustness in modeling baryonic effects compared to the other emulators.

- **BACCO:** BACCO is a neural network-based emulator that accounts for baryonic effects in the non-linear matter power spectrum [12]. BACCO encompasses a parameter set comprising 8 cosmological parameters, consisting of the standard 5 $\Lambda$CDM parameters combined with massive neutrinos and dynamical dark energy. Additionally, it includes 7 free baryonic parameters derived from physical principles, describing factors such as the gas fraction retained in halos, the intensity of AGN feedback, the characteristic galaxy mass, and the relationship between gas fractions and halo mass. In addition to the 7 free parameter model, BACCO also has 3 and 1 parameter models. When not included in the model, the baryonic parameters are fixed at their fiducial values. BACCO achieves an overall precision of $\sim$ 1-5% across its models and its targeted scales ($0.01 < k < 5$ h/Mpc) and redshifts ($0 < z < 1.5$), encompassing various cosmological hydrodynamic simulations. However, BACCO's capacity to confidently predict the baryon-corrected power spectrum is

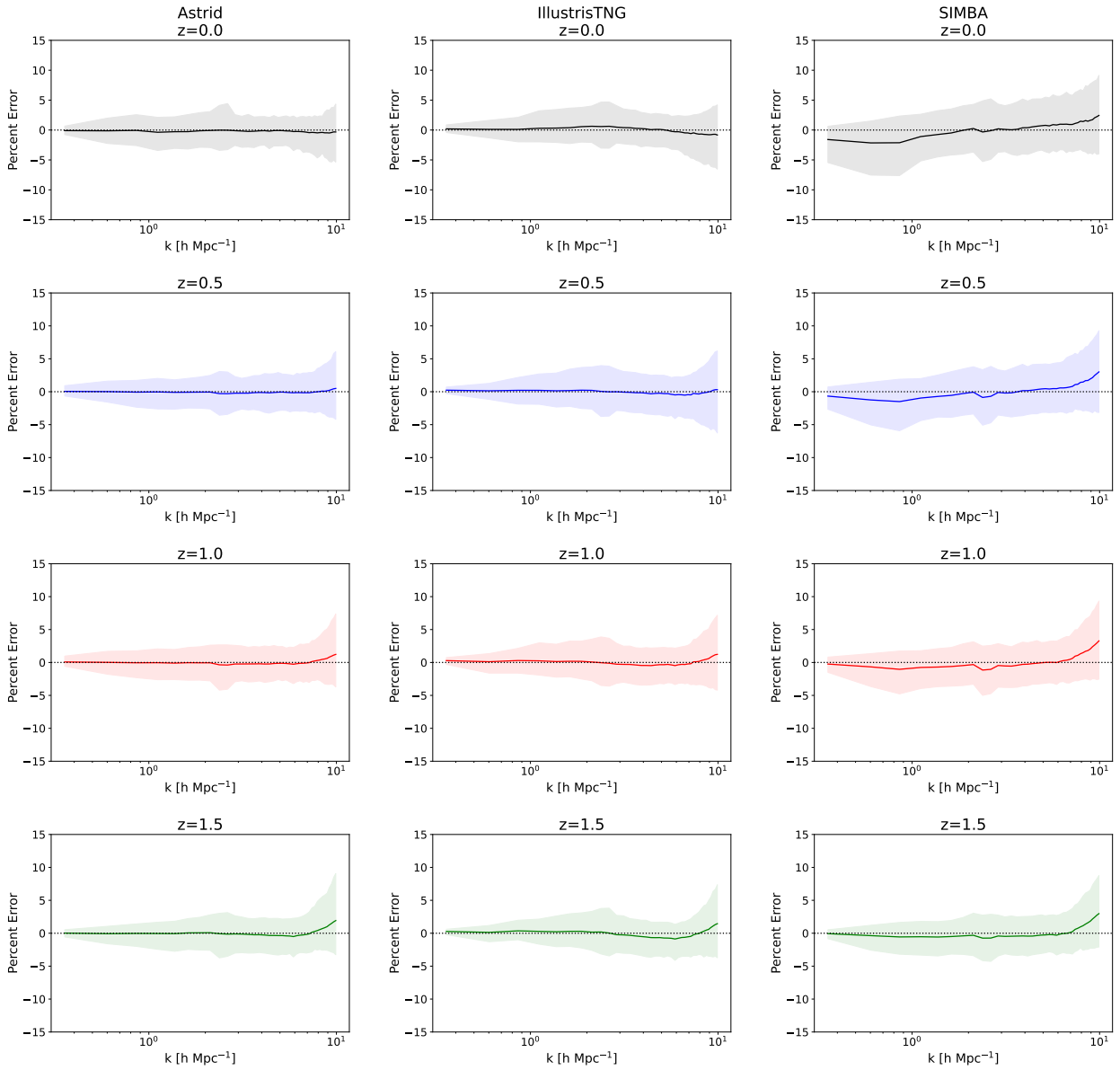



Figure 6.5: Emulator performance on all CAMELS test simulations. The percent error, calculated as $(\text{True Value} - \text{Predicted Value}) \times 100\%$, of emulator predictions is shown as a function of wavenumber. The solid line represents the average percent error across simulations, while the shaded region denotes the 90th percentile range. Notably, the emulator demonstrates exceptional predictive accuracy, with predictions of true baryonic effects consistently achieving accuracy at the percent level. These accuracy results show that the emulator predictions are robust to changes in redshift and hydrodynamic simulation.

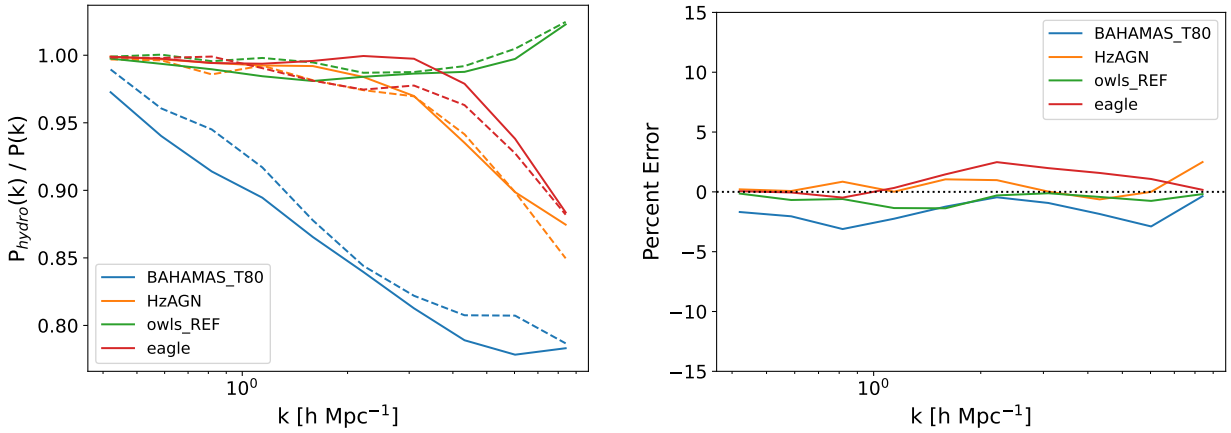



Figure 6.6: The emulator predictions alongside the reference hydrodynamic simulations for scenarios outside the CAMELS dataset. The **left panel** shows the matter power spectrum ratios. Each solid line represents the ground truth from the respective hydrodynamic simulation, while the corresponding dashed line showcases the emulator's predictions. Notably, the emulator closely matches the underlying behavior of the baryonic effects in these distinct simulations. The **right panel** shows the prediction errors for each simulation. The emulator consistently achieves accuracy at the percent level, accurately capturing the baryonic effects in these diverse scenarios. The predictions demonstrate robustness even amidst variations in hydrodynamic simulations, highlighting the emulator's capability to adapt and predict the impact of baryonic physics with high accuracy across a broad spectrum of simulations outside the CAMELS dataset.

limited to a maximum wavenumber of k = 4.7 h/Mpc, notably smaller than our emulator's range. Furthermore, its range of validity is narrower than GPemu: $\sigma_8 \in [0.73, 0.9]$ and $\Omega_m \in [0.23, 0.4]$. As a result, only 39 out of the 200 Astrid $z = 0.0$ test simulations are within BACCO's specified cosmology range.

The left panel of Figure 6.7 shows the comparison between BACCO's predictions (including the 7, 3, and 1 parameter models) and our emulator's predictions on these limited 39 simulations. The solid lines represent the average percent error, while the shaded regions depict the 90th percentile of errors. The dash-dotted and dotted lines illustrate the 90th percentile outputs for BACCO's 3 and 1 parameter emulators, respectively. The comparison results for the SIMBA and IllustrisTNG suites are similar. In Figure 6.8 we compare GPemu against BACCO for the non-CAMELS simulations.

Overall, we find that GPemu exhibits an accuracy similar to that of BACCO, but its range of validity, both in terms of scales and parameter-space, is wider.

- **HMcode:** The HMcode [219] is a simple halo model designed to simulate the influence of baryonic feedback on the power spectrum. It incorporates a six-parameter physical framework

that includes gas expulsion by AGN feedback and encapsulates star formation. The feedback model was fitted to simulation data, taken from the library of [327].

In our evaluation, similar to the comparison conducted against BACCO, we conducted a side-by-side analysis of HMcode's predictions alongside our emulator's outcomes using Astrid test data. The results of this comparison are illustrated in the middle panel of Figure 6.7, with solid lines representing the average percent error and shaded regions depicting the 90th percentile of the errors. We can see that our emulator demonstrates comparable performance to HMcode on larger scales while exhibiting higher accuracy on smaller scales where baryonic effects are stronger. A similar conclusion can be reached by comparing HMCode against GPemu for non-CAMELS simulations as shown in Figure 6.8.

- **BCemu:** The BCemu emulator [104] focuses on modeling the baryonic suppression of the matter power spectrum. It is based on a slightly modified version of the baryonification model [283] and features seven physically-meaningful free-parameters related to gas profiles and stellar abundances within halos. BCemu demonstrated its capability to replicate the power spectra of hydrodynamical simulations with sub-percent precision. Moreover, it established a correlation between the baryonic suppression of the power spectrum and the gas and stellar fractions within halos. However, similar to BACCO, BCemu is constrained by its limited acceptance range for cosmological parameters ($\Omega_m \in [0.196, 0.49]$), encompassing only 148 out of the 200 Astrid test simulations.

  The right panel of Figure 6.7 compares BCemu's predictions with those of our emulator within this subset, with solid lines representing the average percent error and shaded regions depicting the 90th percentile of the errors. From Figure 6.8 we can see that GPemu performs similarly to BCemu when used on non-CAMELS simulations. While both emulators display comparable performance at all scales, our GP emulator shows greater flexibility and generality in its predictions of baryonic effects across a wider range of hydrodynamic simulations.

## Field-level emulation

From Figure 6.4 we found that baryonic effects do not significantly affect the phases of Fourier modes down to $k \sim 10\ h\mathrm{Mpc}^{-1}$. Thus, baryonic effects at the field level can be accounted for by correcting the amplitude of Fourier modes from N-body simulations. Now that we have an emulator for the ratio, $S(k) = P_{\mathrm{hydro}}(k)/P_{\mathrm{nbody}}(k)$, we can investigate how well our model performs at the field level. The resulting field-level transformations exhibit effective improvements, evident across multiple simulation suites at different redshifts.

In more detail, the procedure we employ to model baryonic effects at the field level is as follows. First, we take a given hydrodynamic simulation and its N-body counterpart. We then compute the power spectrum of each of them to compute the baryonic suppression: $S(k) = P_{\mathrm{hydro}}(k)/P_{\mathrm{nbody}}(k)$. Next, we fit the four free parameters of GPemu to get the best match to $S(k)$. Then, from the N-body simulation, we compute the matter density field $\delta_{\mathrm{nbody}}(\mathbf{x})$ and its Fourier transform:

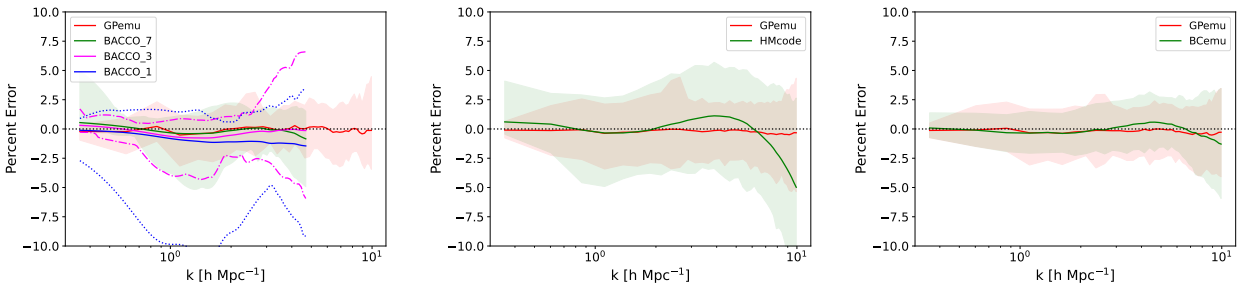



Figure 6.7: Comparison between our GP emulator and other personification emulators - BACCO (left panel), HMcode (middle panel), and BCemu (right panel) - on Astrid test data at $z = 0.0$. The solid lines represent the average percent error, while the shaded regions depict the 90th percentile of errors. **Left Panel:** Within BACCO's acceptable cosmological range, we use 39 test simulations for emulation, covering up to its maximum wavenumber $k \sim 5$. The dash-dotted and dotted lines illustrate the 90th percentile outputs for BACCO's 3 and 1 parameter emulators, respectively. Our GP emulator demonstrates the capability to investigate smaller scales, exhibiting accuracy comparable to BACCO's 7-parameter model while offering increased flexibility and generality. **Middle Panel:** Using all 200 test simulations, our GP emulator demonstrates comparable performance to HMcode on larger scales and exhibits higher accuracy on smaller scales. **Right Panel:** Using the 148 test simulations that fall within BCemu's acceptable cosmological range, both emulators exhibit comparable performance, but our GP emulator showcases greater flexibility and generality in its predictions.

$\delta_{\rm nbody}(\mathbf{k}) = A_{\mathbf{k}} e^{i\theta_{\mathbf{k}}}$. Finally, we obtain the baryon-corrected field by Fourier transforming back $\delta_{\rm postTF}(\mathbf{k})$, where

$$\delta_{\rm postTF}(\mathbf{k}) = \sqrt{S_{\rm GPemu}(k)}\,\delta_{\rm nbody}(\mathbf{k}) \tag{6.14}$$

with $\sqrt{S_{\rm GPemu}(k)}$ being the transfer function predicted by our emulator GPemu.

Figure 6.9 illustrates the baryonic correction of our method on IllustrisTNG when applied to N-body simulations across various redshifts. The first row displays a 2D projection of the whole 3D matter field with dimensions $25 \times 25 \times 25\ (h^{-1}{\rm Mpc})^3$ from a hydrodynamic simulation at four different redshifts. The second row shows the difference between the image from the hydrodynamic simulation and its N-body counterpart. The third row shows instead the difference between the hydrodynamic simulation and our field-level correction model. As expected, our field-level correction is more accurate than the N-body simulation, and the residual fluctuations (shown in red and blue) are due to small-scale modes where the cross-correlation coefficient deviates from 1.

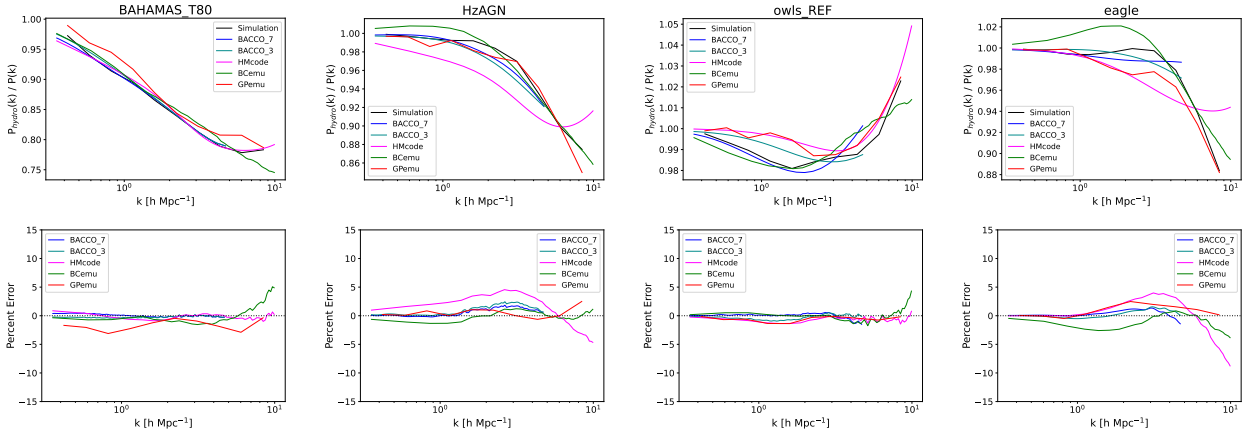



Figure 6.8: Comparison of our emulator against BACCO's 7 and 3 parameter versions, HMcode, and BCemu using hydrodynamic simulations outside of CAMELS. The simulations encompass varying degrees of baryonic feedback. For BACCO, emulation is performed up to BACCO's maximum wavenumber of $k \sim 5$. Despite differences in the length scales of emulation, all emulators achieve percent-level accuracy in predicting baryonic effects, showcasing comparable performance within their respective length scales.

## 6.6 Conclusions

Field-level approaches have the potential to extract all the available information from cosmological surveys. Modeling and marginalizing over baryonic effects at the field-level becomes a key ingredient in these efforts. In this work we have developed a new method to model baryonic effects for the total matter density field, the relevant quantity for weak lensing analyses.

The key finding in this work is that by computing the cross-correlation between the total matter density field in hydrodynamic and N-body simulations from thousands of simulations of the CAMELS project (see Figure 6.4) we conclude that baryonic effects weakly affect the phases of Fourier modes of the total matter density field down to scales as small as $k \sim 10\ h\mathrm{Mpc}^{-1}$. This finding implies that baryonic effects will predominantly modify Fourier mode amplitudes. Thus, we can *baryonify* the total matter field of an N-body simulation by rescaling its Fourier mode amplitudes.

In this work we have built an emulator using Gaussian processes for the total to dark matter power spectrum ratio $S(k)$ that takes as input 2 cosmological parameters ($\Omega_\mathrm{m}$ and $\sigma_8$) and 4 astrophysical parameters ($A_\mathrm{SN1}$, $A_\mathrm{SN2}$, $A_\mathrm{AGN1}$, $A_\mathrm{AGN2}$). We have trained our emulator using 800 state-of-the-art hydrodynamic simulations from the Astrid suite of CAMELS. We then show that our emulator is able to reproduce the baryonic effects of thousands of hydrodynamic simulations that have different cosmologies, astrophysics, subgrid physics, resolutions, volumes, and redshifts within a few percent precision.

We have compared our emulator against others in the literature, such as BACCO, HMCode,

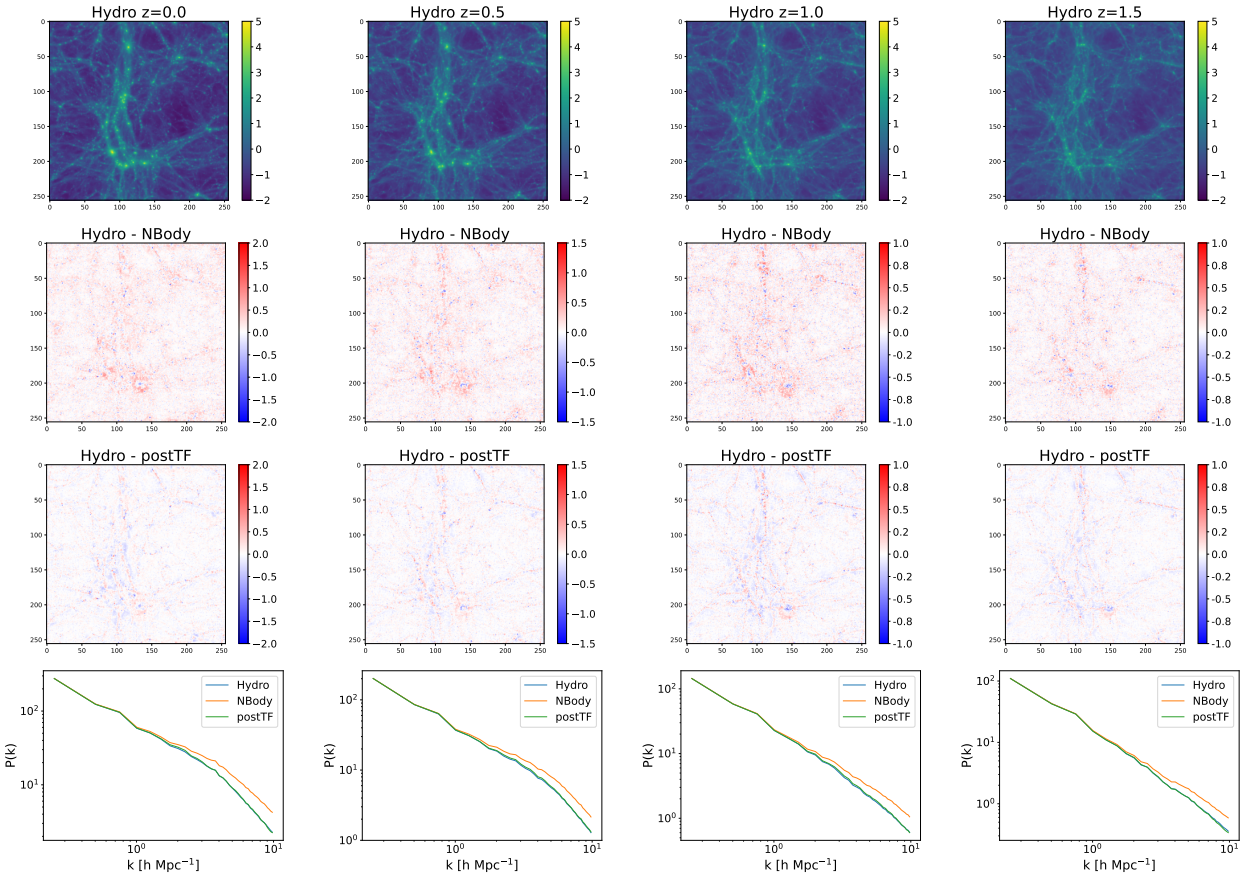



Figure 6.9: Field-level improvement comparison in the IllustisTNG simulation suite across different redshifts. The top row presents the original hydrodynamic field, while the bottom row shows the power spectra of the resultant N-body field after applying the transfer function, showcasing a substantial enhancement in power spectra alignment with the hydrodynamic fields. The achieved near-perfect calibration signifies improved field-level agreement, based on the implications from figures 6.1 and 6.4. The second row displays the residual differences between hydrodynamic and initial N-body fields, whereas the third row demonstrates the residuals post-application of the transfer function (postTF). Notably, the post-transfer function residuals reveal a predominantly white background, indicating significantly improved agreement on both large scales and around clustered regions and halos on smaller scales. This illustrates the transfer function's efficacy in achieving notable field-level improvements, showcasing robustness across various redshifts.

and BCemu. We find that our emulator shares a similar level of accuracy with those, but it has a wider range of validity given that it has been trained on CAMELS, where variations in cosmology and astrophysics are very large. We also showed explicitly how using our method reduces the residuals when working at the field level by comparing the results of hydrodynamic simulations against *baryonified* N-body simulations. A limitation of using CAMELS is that the box size is very small, and baryonic effects may not be fully captured due to the absence of larger halos in these simulation boxes. This will need to be investigated in more detail using a suite of simulations varying box size.

Our emulator enables robust, cost-effective field-level weak lensing modeling and facilitates precise power spectra analyses at the two-point level. The versatility and accuracy of our GP baryonification emulator underscore its potential as a powerful tool in cosmological simulations, offering opportunities for enhanced analyses and deeper insights into baryonic effects in large-scale structures. However, whether this emulator suffices at the field level depends on the specifics of the observational program. For example, for weak lensing, this will require making weak lensing maps using ray-tracing techniques. The overall detectability of the effects that go beyond our field level emulator in the weak lensing depends on the density of background galaxies and the observed area of the sky. This analysis goes beyond the purpose of this paper, and will be presented elsewhere.

# Acknowledgements

We thank Raul Angulo and Aurel Schneider for their comments on the usage of the BACCO and BCemu emulators. DS thanks James Sullivan for helpful discussions on Gaussian processes. The work of FVN is supported by the Simons Foundation. The CAMELS project is supported by the Simons Foundation and the NSF grant AST 2108078.

# Chapter 7

# Translation and Rotation Equivariant Normalizing Flow (TRENF) for Optimal Cosmological Analysis

The contents of this chapter were originally published in [56],

Translation and Rotation Equivariant Normalizing Flow (TRENF) for Optimal Cosmological Analysis

*Dai B., Seljak U.*(arXiv:2202.05282) MNRAS 516.2 (2022): 2363-2373

Our universe is homogeneous and isotropic, and its perturbations obey translation and rotation symmetry. In this work we develop Translation and Rotation Equivariant Normalizing Flow (TRENF), a generative Normalizing Flow (NF) model which explicitly incorporates these symmetries, defining the data likelihood via a sequence of Fourier space-based convolutions and pixel-wise nonlinear transforms. TRENF gives direct access to the high dimensional data likelihood $p(x|y)$ as a function of the labels $y$, such as cosmological parameters. In contrast to traditional analyses based on summary statistics, the NF approach has no loss of information since it preserves the full dimensionality of the data. On Gaussian random fields, the TRENF likelihood agrees well with the analytical expression and saturates the Fisher information content in the labels $y$. On nonlinear cosmological overdensity fields from N-body simulations, TRENF leads to significant improvements in constraining power over the standard power spectrum summary statistic. TRENF is also a generative model of the data, and we show that TRENF samples agree well with the N-body simulations it trained on, and that the inverse mapping of the data agrees well with a Gaussian white noise both visually and on various summary statistics: when this is perfectly achieved the resulting $p(x|y)$ likelihood analysis becomes optimal. Finally, we develop a generalization of this model that can handle effects that break the symmetry of the data, such as the survey mask, which enables likelihood analysis on data without periodic boundaries.

## 7.1 Introduction

The goal of optimal cosmological analysis is to extract the maximum amount of information of cosmological parameters from the data. If the data are Gaussian distributed this task has a well-known solution, as one can directly evaluate the Gaussian data likelihood $p(x|y)$, where $y$ are cosmological parameters of interest. An implementation of this method is the optimal quadratic estimator [117, 315, 30], which uses second-order expansion of the likelihood to achieve this task. Often we must also model the likelihood as a function of nuisance parameters such as systematics, astrophysical sources, etc. Evaluating either the quadratic estimator or the likelihood in high dimensions is not an easy task in the presence of noise and survey mask [108], since it requires an inversion and a determinant of the covariance matrix, which for large data is prohibitively expensive. Often simplified versions such as the pseudo power spectrum analysis are adopted [133]. These however are suboptimal and may not be amenable to a fast evaluation of covariance matrix, which for optimal quadratic estimator is available in the form of the Fisher matrix.

In the nonlinear regime, such as the large -scale structure on small scales, nonlinear gravitational effects in dark matter create a rapidly growing cascade of higher-order correlations, which quickly get populated at all orders. In this regime we often rely on N-body simulations. Furthermore, what we often observe are baryons, such as galaxy light or gas density, which have additional astrophysical processes that need to be included in the simulation. These are handled with nuisance parameters that try to parametrize the unknown physics.

When it comes to data analysis in this regime the exact likelihood analysis is deemed impossible, and instead the focus has been on extracting information from a limited set of summary statistics. This program has numerous challenges. The first is how to choose the most informative summary statistics. While two-point function is a natural choice even in the nonlinear regime, adding higher -order information is less straightforward. Even adding the three-point function means adding a function of three parameters, which is considerably more complex to describe than the two-point function or its Fourier analog, the power spectrum. Numerous other ad-hoc statistics $S(x)$ have been proposed, from peak counts to void counts, void profiles, etc. In each case, one must evaluate their mean expectation as a function of cosmological and nuisance parameters. Moreover, since these are ad-hoc summary statistics their probability distribution must be evaluated numerically, usually by approximating the probability distribution as a multi-variate Gaussian. Recent developments such as Likelihood Free Inference [7, 8] or Simulation Based Inference [51] pursue this program by combining the two tasks of mean and covariance estimate into estimating the full $p(S|y)$, which can also include effects beyond the mean and covariance of the Gaussian distribution. These have the same underlying issues of summary statistics being ad-hoc and potentially sub-optimal. The summary statistics can also be determined by some information maximizing Machine Learning algorithms [41, 270, 202].

An alternative approach is using the reconstruction of initial conditions and performing the marginal integral over these latent variables so that we are left with marginal data likelihood as a function of cosmological parameters $p(x|y)$. However, performing this marginal integral over initial conditions $z$ has proven to be difficult. One can sample over $z$ and $y$ using Hamiltonian Monte Carlo (HMC) [147, 163, 346], but the samples are usually very correlated in high dimensions

even if HMC is used, and thousands of full N-body simulation steps may be needed between two independent samples. An alternative to this approach is maximum a posteriori (MAP) estimation of $z$, followed by an MAP-based unbiased estimator of the cosmological parameters [287, 224]. While obtaining the MAP of $z$ is faster, making the estimate unbiased and obtaining the cosmological parameter posterior is still expensive and can be suboptimal.

In this paper we propose instead to learn directly the data likelihood $p(x|y)$ from the data simulations conditioned on $y$. Since the data is very high dimensional, and the simulations are expensive, this task has been deemed difficult or impossible. However, the universe is homogeneous and isotropic on average, and its perturbations obey translational and rotational symmetry in a statistical sense. If these symmetries can be imposed into the structure of $p(x|y)$, the parametrization of the model would be greatly restricted, allowing efficient learning of $p(x|y)$. To see the power of symmetries we consider a simple example of modeling an N-dimensional Gaussian Random Field (GRF), where one needs $N(N+3)/2$ parameters to describe its mean and covariance matrix. However, if the GRF satisfies translation and rotation symmetry, the mean vector is reduced to a scalar, and the covariance matrix is reduced to a 1D function, i.e., the power spectrum, which can usually be parameterized by only a few parameters assuming smoothness. This simple example shows that symmetries can greatly reduce the degrees of freedom of the model and the sample complexity. In this work we will use the framework of generative learning to learn the data likelihood $p(x|y)$ and build the symmetries into the model itself.

Latent variable generative models such as Normalizing Flows (NFs) [268, 70, 71, 161, 57], Variational Auto-Encoders (VAEs) [162, 269] and Generative Adversarial Networks (GANs) [107, 262] aim to model the high-dimensional data distribution $p(x)$ by introducing a mapping from a latent variable $z$ to $x$, where $z$ is assumed to follow a given prior distribution $\pi(z)$. While all these three classes of models have been shown to produce realistic samples [161, 267, 157], NF is the only one that allows exact density evaluation $p(x)$, and when done conditionally as $p(x|y)$ this enables a direct likelihood analysis. Another family of density estimation models is called autoregressive models [103, 238], which decompose the high dimensional Probability Distribution Function (PDF) as the product of 1D conditional PDFs: $p(x) = \prod_{i=1}^{N} p(x_i|x_{1:i-1})$. These models require choosing a specific ordering of the pixels $x_{1:N}$ and treat the pixels differently, making it hard to enforce symmetries. For these reasons we will adopt NF as the method of choice for cosmological data analysis.

Translation symmetry can be implemented by Convolutional Neural Networks, and in cosmology these have been used for example to relate galaxies to dark matter [224, 42]. In machine learning the symmetries other than translation are often included using brute force methods such as data augmentation. This increases the amount of training data and does not reduce the complexity of the model, which must instead learn the symmetries from the data samples. There is also no guarantee that the symmetries can be perfectly learned. In such situations NFs can fail in their primary tasks, either as a realistic data generator or as a data likelihood estimator. There are also works trying to build the symmetries into the machine learning models [46, 348, 355, 347, 218], but these models are mostly designed for supervised tasks such as classification [277] and high dimensional mapping, and cannot be directly used in the NF framework, which requires the learned mapping to be invertible and to have tractable Jacobian determinant. NFs have been applied to

modeling cosmological fields [274, 121] for data generation, but in high dimensions, good sample qualities do not necessarily imply accurate likelihood estimation [318], which is the primary goal of this paper. Symmetries are also not incorporated in these models. In this work we develop Translation and Rotation Equivariant Normalizing Flow (TRENF), which impose the symmetries explicitly into the NF model.

The novel developments of this paper are:

- We develop a conditional NF architecture which is translation and rotation equivariant (TRENF) for learning the likelihood $p(x|y)$ of cosmological fields.
- We use TRENF as a map from latent space to data space for fast generation of high dimensional simulated data conditional on cosmological parameters. Note that cosmological fields (e.g., Cosmic Microwave Background) are usually high dimensional distributions and cannot be approximated by low dimensional manifolds, and therefore modeling them with low-dimensional-manifold models like GANs could potentially introduce systematics in the samples. TRENF, on the other hand, has no dimension reduction and can sample from the full distribution $p(x|y)$ without any manifold assumption.
- We use TRENF as a map from the data space to the latent space, enabling visual and numerical inspection of the quality of the training: if the latent map is a perfect white noise Gaussian at the true value of $y$, then we have optimally extracted all the information from the data, encoding it into a single number $p(x|y)$. TRENF thus can identify when the model is incomplete, such as missing some systematic or modeling effect. Recent works on applying CNNs [270] or novel summary statistics [44] to extract information from the fields have shown improvements over traditional summary statistics like power spectrum, but are not providing any guarantees of optimality, and it is unclear how much information has remained unused. Our generative model provides a natural way to investigate this and improve upon these methods.
- We use TRENF $p(x|y)$ as a function of $y$ to directly provide uncertainty quantification via the posterior $p(y|x) = p(x|y)p(y)/p(x)$, which is the ultimate goal of a cosmological analysis.
- We introduce non-symmetric components into TRENF for modeling observational effects that break the symmetry of the data, such as the survey mask.

One limitation of the current work is that it is limited to two-dimensional fields. While we will be presenting results on projected dark matter fields, there are many other projections available observationally, such as weak lensing convergence, projected galaxy density fields, etc. In fact, most of the large-scale structure data are 2D projections. Moreover, for 2D fields, most of the complicating effects such as light-cone projections, baryonic effects, and redshift space distortions, do not break translation and rotation equivariance. One effect that does break is the survey mask, and we address it in detail in this paper.

## 7.2 Method

### Normalizing Flows

Flow-based models provide a powerful framework for density estimation [71, 239] and sampling [161]. These models map the data $x$ to latent variables $z$ through a sequence of invertible transformations $f = f_1 \circ f_2 \circ ... \circ f_n$, such that $z = f(x)$ and $z$ is mapped to a base distribution $\pi(z)$. The base distribution $\pi(z)$ is normally chosen to be a standard normal distribution, i.e. a Gaussian white noise with zero mean and unit variance, $\pi(z) = \mathcal{N}(0, I)$. The probability density of data $x$ can be evaluated using the change of variables formula:

$$\begin{aligned} p(x) = & \quad \pi(f(x)) \left|\det\left(\frac{\partial f(x)}{\partial x}\right)\right| \\ = & \quad \pi(f(x)) \prod_{l=1}^{n} \left|\det\left(\frac{\partial f_l(x)}{\partial x}\right)\right|. \end{aligned} \tag{7.1}$$

To sample from $p(x)$, one first samples latent variable $z$ from $\pi(z)$, and then transform variable $z$ to $x$ through $x = f^{-1}(z)$. The transformation $f$ is usually parametrized with neural networks $f_\phi$, and the parameters $\phi$ are estimated using Maximum Likelihood Estimation (MLE):

$$\phi^* = \arg\max_{\phi} \frac{1}{N} \sum_{i=1}^{N} \log p_\phi(x_i), \tag{7.2}$$

where the data likelihood $p(x)$ is given by Equation 7.1. The MLE solution minimizes the Kullback-Leibler (KL) divergence between the model distribution $p_\phi(x)$ and the true data distribution. The parametrization of $f$ must satisfy the requirements that the Jacobian determinant $\det(\frac{\partial f_l(x)}{\partial x})$ is easy to compute for evaluating the density, and the transformation $f_l$ is easy to invert for efficient sampling.

### Translation and Rotation Symmetry

It is useful to differentiate the concepts invariant and equivariant. A function $f$ is invariant if its output is unchanged when its input $x$ is transformed by a symmetry group $g$:

$$f(g \cdot x) = f(x). \tag{7.3}$$

A relevant example is the PDF of the cosmological fields, which should be invariant under translation and rotation of the fields. Similarly, a function $f$ is equivariant if its output is transformed by the same symmetry group $g$ as its input $x$:

$$f(g \cdot x) = g \cdot f(x). \tag{7.4}$$

In other words, an equivariant function commutes with the symmetry transformation. The physical laws that govern the evolution of our universe are equivariant to translation and rotation, if we view them as a mapping from the early universe to the present day. We want our NF transformation $f$ to have similar properties as the physical laws, thus to be equivariant to these symmetries. An equivariant NF $f$ also leads to invariant PDF from Equation 7.1.

## Translation and Rotation Equivariant Normalizing Flow (TRENF)

Our goal is to find a parametrization of the flow transformation $f$ such that 1) its Jacobian determinant and inverse can be efficiently calculated for likelihood evaluation (Equation 7.1) and sampling; 2) $f$ is equivariant to translation and rotation. The simplest form of such transformation is the Pixelwise Gaussianization (PG), which applies the same nonlinear transformation on all pixels such that the resulting one-point PDF is a standard Gaussian. This method has been used to reconstruct the primordial density fields [350]. However, PG is not very expressive and cannot model the correlations between different pixels. Here we discuss how to go beyond PG.

We observe that for any non-Gaussian PDF that is invariant to translation, one can always find a convolution kernel, such that the one-point PDF of the convolved field is non-Gaussian. This can be proven by considering the non-zero high-order cumulant of the field. Suppose the m-point cumulant is non-zero $\langle x_1 x_2 \cdots x_m \rangle_c \neq 0$ ($m > 2$). We can always define a convolution kernel $T$ that is nonzero at $x_1, x_2, \cdots x_m$. The m-th cumulant of the one-point PDF of the convolved field contains $\langle x_1 x_2 \cdots x_m \rangle_c$ and must be nonzero for some kernel $T$.

This motivates parametrizing the flow transformation $f$ with convolutions followed by PG. Intuitively, the convolution kernels look for maximal non-Gaussianity in convolved data, which indicates non-Gaussian PDF and non-zero high-order cumulants (order above two). The PG maps the one-point PDF of the convolved data to a Gaussian and reduces the high-order cumulants. By stacking multiple such transformations, all high-order cumulants can be reduced to zero and the data distribution is mapped to a white noise Gaussian. This process can be viewed as a generalization of SINF (sliced iterative NF) [57] to translation equivariant data: SINF also searches for maximally non-Gaussian directions, followed by PG. For translation equivariant data these directions are replaced with convolutions.

Motivated by these ideas, we choose to parametrize $f$ with convolutions and pixel-wise non-linearity. Assuming periodic condition, the convolution of data $x(\boldsymbol{r})$ can be written in Fourier space as

$$\int T(\boldsymbol{r} - \boldsymbol{r'}) x(\boldsymbol{r'})\, d\boldsymbol{r'} = \hat{F}^{-1}\left(\tilde{T}(k) \cdot \tilde{x}(\boldsymbol{k})\right) \tag{7.5}$$

where $T$ is the convolution kernel, $\hat{F}$ denotes Fourier transform, $\tilde{T} = \hat{F}(T)$ and $\tilde{x} = \hat{F}(x)$ are the Fourier transform of $T$ and $x$, respectively. We require the convolution to be rotational equivariant, thus $\tilde{T}$ can only depend on $k$, the amplitude of $\boldsymbol{k}$. We combine a convolution operation with a pixelwise non-linearity $\Psi$ to form the basic transformation of TRENF:

$$f = \Psi\left(\hat{F}^{-1}\tilde{T}(k)\hat{F}x\right). \tag{7.6}$$

Both $\tilde{T}$ and $\Psi$ are 1D functions learned from the data. We choose to parametrize $\tilde{T}(k)$ with cubic splines. $\Psi$ function is required to be monotonic and differentiable in order to sample and evaluate density from TRENF. We parametrize $\Psi$ with monotonic rational quadratic splines [110, 80]. We will refer to each such transformation as one layer in the rest of this paper.

Equation 7.6 satisfies the two requirements we set at the beginning of this section. Firstly, its Jacobian determinant and inverse can be calculated via [156]:

$$\left|\frac{df}{dx}\right| = \left(\prod_{i}^{\text{pixels}} \frac{d\Psi}{dx(\boldsymbol{r}_i)}\right) \cdot \left(\prod_{j}^{\text{k modes}} \tilde{T}(k_j)\right), \tag{7.7}$$

$$f^{-1} = \hat{F}^{-1}(1/\tilde{T})\hat{F}\Psi^{-1}(x). \tag{7.8}$$

Secondly, it can be easily verified that both the convolution and the pixel-wise non-linearity are translational and rotational equivariant, so Equation 7.6 also satisfies the symmetry requirement. To improve the expressivity of the model, one can stack multiple such transformations and form a deep NF model.

The architecture of TRENF is similar to a Convolutional Neural Network (CNN): both of them are composed of convolutions followed by non-linearities. While CNNs normally perform convolution in real space, this becomes too expensive for long range correlations typical of cosmology data. TRENF compute the convolution in Fourier space, making it possible to easily calculate its inverse and Jacobian determinant. The Fourier space also allows us to easily enforce rotational symmetry, and parametrize arbitrarily large kernels. In the NF framework we keep the dimensionality of the data, unlike CNNs which usually change the number of channels and side lengths. Another difference between TRENF and CNN is that the non-linearity is learnable in TRENF, which in CNN it is normally chosen to be a fixed function like ReLU. This extra degree of freedom increases the flexibility of TRENF.

## Conditional TRENF

To learn the model dependence on cosmology, baryonic physics, and other nuisance parameters, we build conditional TRENF where the model parameters are functions of conditional variables $y$. Specifically, we train a hyper neural network $g$ to learn the conditional relation $\phi = g(y)$, where $\phi$ consists of the spline parameters of all the kernel $\tilde{T}$ and non-linearity $\Psi$. The total number of TRENF parameters $\phi$ is usually of order $\mathcal{O}(100)$, and they can be easily predicted by a single fully-connected hyper network. An alternative is to interpolate between different $y$ using a Gaussian Process.

## Training

We explore two kinds of training losses in this work: generative loss and discriminative loss. In the generative loss we minimize the negative log-likelihood, which is the standard loss function of NF (Equation 7.2 with conditional variable y):

$$\mathcal{L}_{\text{g}} = -\frac{1}{N}\sum_{i=1}^{N} \log p(x_i|y_i). \tag{7.9}$$

The generative training is suitable for sampling (Section 7.3) and conditional density estimation.

For posterior analysis in Section 7.4, our task is to obtain the most accurate posterior distribution, rather than the likelihood function. It has also been shown that discriminative learning with objective $p(y|x)$ generally has a lower asymptotic error on classification tasks than generative learning with objective $p(x|y)$ [235]. Therefore, we adopt a two-stage training strategy, where we firstly train TRENF with a generative loss, and then we switch to the discriminative loss ($-\log p(y|x)$) to improve the accuracy of the posteriors. The first stage (generative learning) can be viewed as an initialization (warm startup) of the discriminative learning and speeds up the training process. In the second stage, the calculation of $\log p(y|x)$ involves computing the evidence $p(x)$, which is estimated using Importance Sampling (IS):

$$\begin{aligned}
\mathcal{L}_{\mathrm{d}} &= -\frac{1}{N}\sum_{i=1}^{N}\log p(y_i|x_i) \\
&= -\frac{1}{N}\sum_{i=1}^{N}\left[\log p(x_i|y_i) + \log p(y_i) - \log\left(\int p(x_i|y)p(y)dy\right)\right] \\
&\approx -\frac{1}{N}\sum_{i=1}^{N}\Bigg[\log p(x_i|y_i) + \log p(y_i) \\
&\quad - \log\left(\frac{1}{M}\sum_{y_j\sim q(y|x_i)}^{M}\frac{p(x_i|y_j)p(y_j)}{q(y_j|x_i)}\right)\Bigg].
\end{aligned} \tag{7.10}$$

During training, for each training data $x_i$, we first find the MAP $y_{i,\mathrm{MAP}} = \arg\max p(x_i|y)p(y)$ using ADAM optimizer, and then the IS distribution $q(y|x_i)$ is defined as a Gaussian centered at $y_{i,\mathrm{MAP}}$ with a fixed covariance matrix. The parameters of the covariance matrix are fitted to the posterior distribution of the first stage training. Note that if we skip the first stage training and directly train TRENF with Equation 7.10, the Gaussian $q(y|x_i)$ is normally a poor approximation to the true posterior, and the optimization is difficult to converge due to inaccurate estimation of $p(x)$. The number of importance sampling points $M$ is a hyperparameter, and we use $M = 20$ for the datasets we considered in this paper.

## Modeling Effects that Break the Data Symmetry

The physical fields satisfy translation and rotation symmetries, but our observed data usually do not. There are several observational effects, such as the survey mask and foregrounds, that could break the symmetry of the data. To model these effects we need to introduce non-equivariant components into our model. The idea is that the non-equivariant component will model these observational effects that break the symmetries, while TRENF takes care of the physical process that obeys the symmetries.

As data preprocessing, we first sample Gaussian random noise to the missing pixels $x_{\mathrm{mask}}$ so that the data have rectangular shapes. We then add an affine coupling layer [71] which applies affine

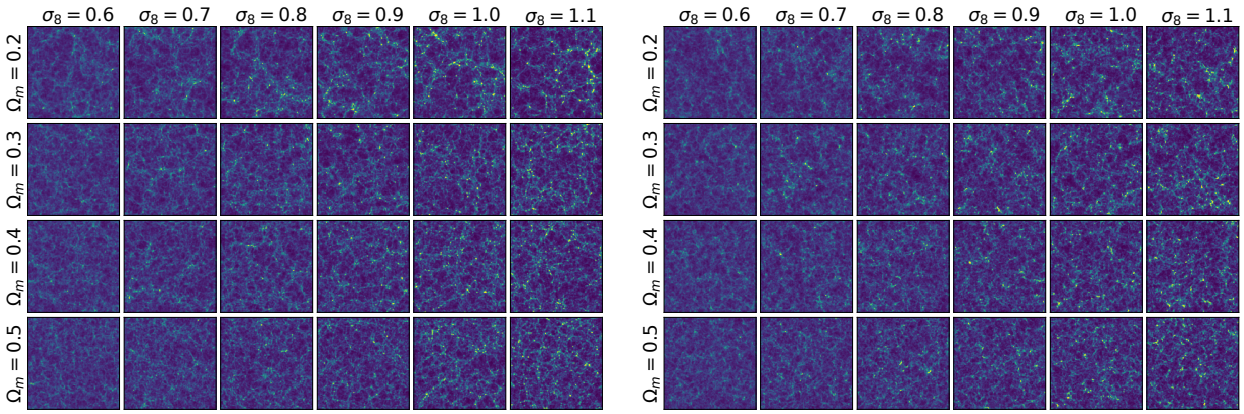



Figure 7.1: Test data (left panel) and uncurated TRENF samples (right panel) as a function of conditional variables $\Omega_m$ and $\sigma_8$.

transforms on $x_{\text{mask}}$, conditional on the neighboring pixels $x_{\text{neighbor}}$ and conditional variables $y$:

$$z = x_{\text{mask}} \odot \exp(s(x_{\text{neighbor}}, y)) + t(x_{\text{neighbor}}, y), \tag{7.11}$$

where $\odot$ denotes element-wise product, and $s$ and $t$ are functions modeled by neural networks. The other pixels are left unchanged in this layer. This step can be effectively seen as inpainting, even though we do not explicitly train the layer to recover the missing pixels, but rather train the whole model using NF objectives.

After the affine coupling layer, we add convolutions and pixel-wise non-linearities similar to Equation 7.6 to map the data to a Gaussian. Here we introduce non-equivariant components into these transforms to model effects like non-periodic boundaries and position-dependent noise. Since these effects are usually position-dependent, in this work we choose to introduce position dependence on the non-linearity $\Psi(x) = \Psi_{\boldsymbol{r}}(x)$. Specifically, we train two separate hyper networks $g_{\tilde{T}}$ and $g_\Psi$. $g_{\tilde{T}}$ models the dependence of convolution kernel parameters $\phi_{\tilde{T}}$ on conditional variable $y$:

$$\phi_{\tilde{T}} = g_{\tilde{T}}(y), \tag{7.12}$$

while $g_\Psi$ models the conditional dependence of non-linearity parameters $\phi_\Psi$ on position $\boldsymbol{r}$ and $y$:

$$\phi_\Psi = g_\Psi(\boldsymbol{r}, y). \tag{7.13}$$

## 7.3 Results: generative samples in data space and data representation in latent space

**Dataset:** The dataset we will use throughout the paper is 2D projections of matter overdensity fields. This example is most relevant for weak lensing applications, which are similar projections of matter density field along the line of sight. Here, for the initial analysis, we want to have the

data to be periodic, so that translation and rotation symmetry is not broken. In section 7.5 we will generalize it to non-periodic boundaries.

The matter overdensity fields are generated by N-body solver FastPM [89]. We uniformly sample $\Omega_m$ and $\sigma_8$ from the prior $\Omega_m \in [0.2, 0.5]$ and $\sigma_8 \in [0.5, 1.1]$, and fix the other cosmological parameters to Planck 2015 [256]. $\Omega_m$ and $\sigma_8$ are the conditional variable $y$ in this study. Each simulation is run in a 512 $h^{-1}\mathrm{Mpc}$ box with 10 time steps using $128^3$ particles. The matter overdensity field at redshift $0$ is measured on a $128^3$ mesh. Then we divide the box into four slices (128 $h^{-1}\mathrm{Mpc}$ thick) along z-axis, and project each slice along z-axis to get four $128^2$ matter overdensity fields.

**TRENF:** We build a TRENF model consisting of 5 transformation blocks (Equation 7.6), with $8$ spline points in $\tilde{T}$ and $\Psi$ in each block. The hyper network is chosen to be a multilayer perceptron with $2$ hidden layers and $512$ neurons in each hidden layer. Since our input data $1 + \delta$ is non-negative, the data preprocessing is performed by first removing the $[0, +\infty)$ boundary with an inverse softplus transform

$$\mathrm{invsoftplus}(1 + \delta) = \log(e^{1+\delta} + \epsilon - 1), \tag{7.14}$$

followed by a normalization layer $\frac{x-\mu}{\sigma}$ to scale the data to zero mean and unity variance. Here $\mu$ and $\sigma$ are both scalars that are independent of conditional variable $y$. The generative loss function (Equation 7.9) is used to optimize the model in the first stage.

## Generative samples in data space

Once we have trained the NF we can draw a vector $z$ from a white noise distribution and map it into the data space via $x = f_y^{-1}(z)$. Figure 7.1 shows the resulting maps sampled from the trained TRENF, comparing them to the test data. The training and sampling are conditioned on cosmological parameters $\Omega_m$ and $\sigma_8$. We see that the samples have a similar structure as the test data, and reproduce the nonlinear evolution of structure with $\sigma_8$ (structure becoming more nonlinear with $\sigma_8$), and voids becoming smaller with $\Omega_m$.

In Figure 7.2 we show various statistics run on test data and on TRENF samples. We compare them in terms of the power spectrum, one-point distribution function at the pixel scale, and the bispectrum. In all cases the agreement is nearly perfect, suggesting that TRENF samples are not only visually correct but also reproduce the low and high order statistics.

TRENF takes about $0.4$ second to generate 100 images on a Tesla V100 GPU. The simulations we trained on are computationally cheap, as they are generated with fast Particle-Mesh simulations FastPM with only 10 time steps, so the computing time is about $40$ seconds on a CPU. One could also train TRENF on output maps obtained from full N-body simulations or hydrodynamical simulations, and the computational gain in sampling time of a simulation image would be more significant. This shows the promise of TRENF as a fast and realistic generative model for cosmological data such as galaxy images, weak lensing maps, Sunyaev-Zeldovich maps (thermal and kinetic), etc.

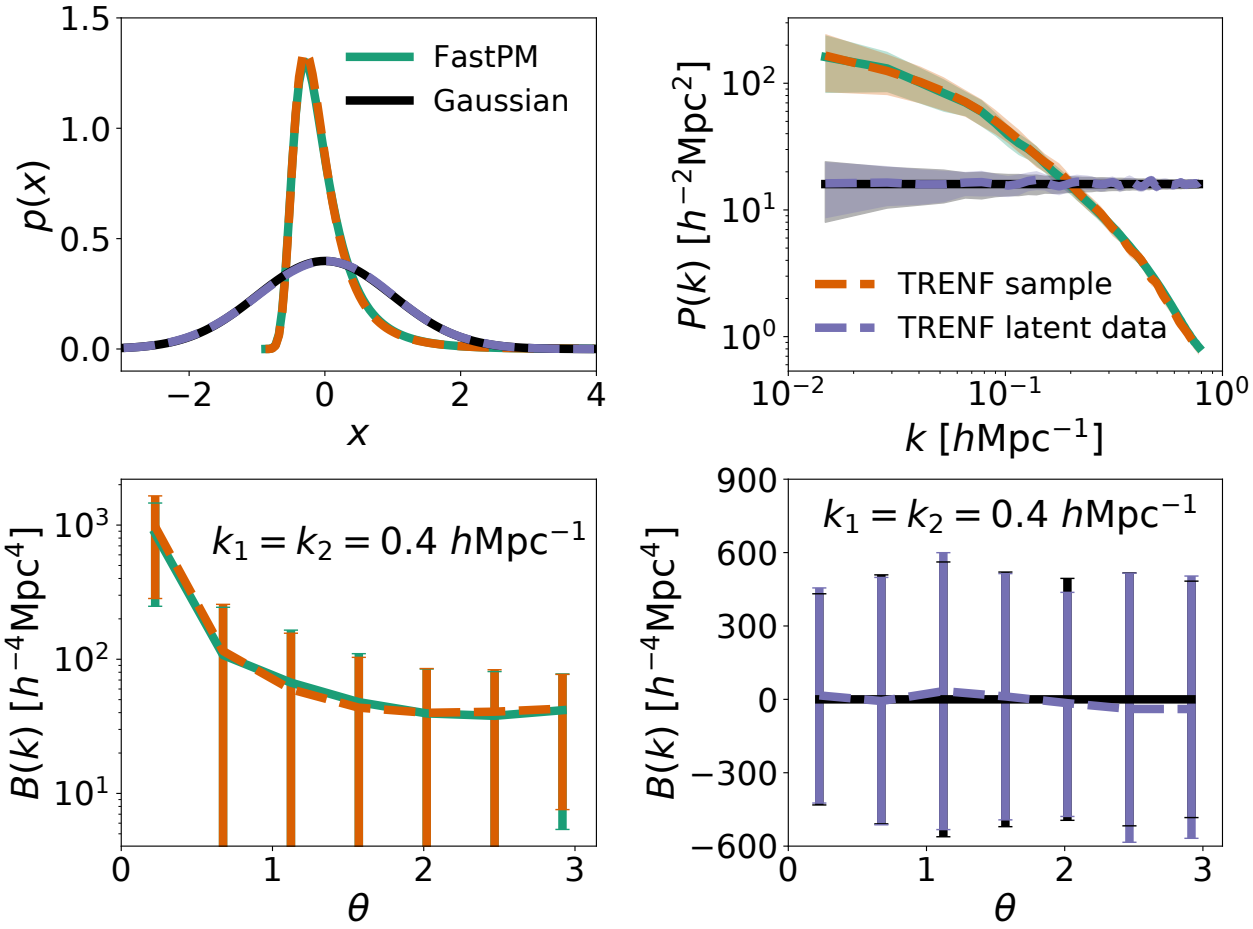



Figure 7.2: The pixel probability distribution function (upper left), power spectrum (upper right), and bispectrum (lower panels) of FastPM, TRENF samples, and FastPM data in TRENF latent space. All the results are measured over 10000 samples. The shadowed regions in the power spectrum plot and the error bars in the bispectrum plot indicate $16\%$ and $84\%$ of the distribution. The samples of TRENF agree well with FastPM on these summary statistics. In TRENF latent space FastPM data is consistent with Gaussian white noise.

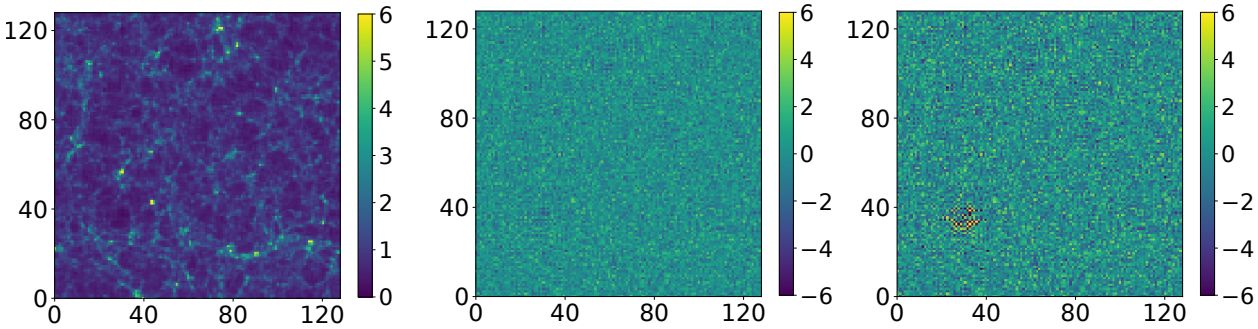



Figure 7.3: Test data (left panel), latent data transformed with correct cosmology ($\Omega_m = 0.309$, $\sigma_8 = 0.816$, middle panel), and latent data transformed with incorrect cosmology ($\Omega_m = 0.4$, $\sigma_8 = 0.5$, right panel).

### Data representation in latent space

The training of TRENF achieves its goal of optimal likelihood if it maps the data into the target distribution, for which we use a Gaussian white noise distribution. To test this we compare the test data mapped into the latent space with the standard Gaussian distribution. In Figure 7.3 we show the visualization of the latent data transformed with the correct conditional variable $y$ and incorrect $y$. We can see that when the correct $y$ is used the latent space is visually indistinguishable from Gaussian white noise. We show their one-point PDFs, power spectra and bispectra in Figure 7.2. On all these summary statistics the latent data are consistent with the standard Gaussian. This is very encouraging: if the latent data distribution is a perfect Gaussian white noise, then we have achieved optimal NF, and the resulting $p(x|y)$ contains all the information of the data $x$. In contrast, when we use an incorrect $y$ the map is no longer white Gaussian. For example, we see strange patterns in the latent map at the position of a large void in the data space (Figure 7.3).

## 7.4 Results: likelihood and posterior analysis

From the perspective of optimal cosmological analysis, the most powerful component of NFs is their ability to provide conditional density or likelihood $p(x|y)$. If the likelihood is extracted optimally then we can achieve optimal cosmological analysis. To establish the ability of TRENF to extract the likelihood we turn first to a Gaussian Random Field (GRF) example, where the information content of the data and the likelihood of the data are both known analytically.

### Gaussian Random Fields

Similar to the matter overdensity map, we generate GRFs $\delta(r)$ in 512 $h^{-1}\mathrm{Mpc}$ boxes with $128^2$ resolution. The halofit power spectrum [310] at redshift 0 is used to generate the data with cosmological parameters $\Omega_m$ and $\sigma_8$ uniformly sampled from the same range $\Omega_m \in [0.2, 0.5]$ and

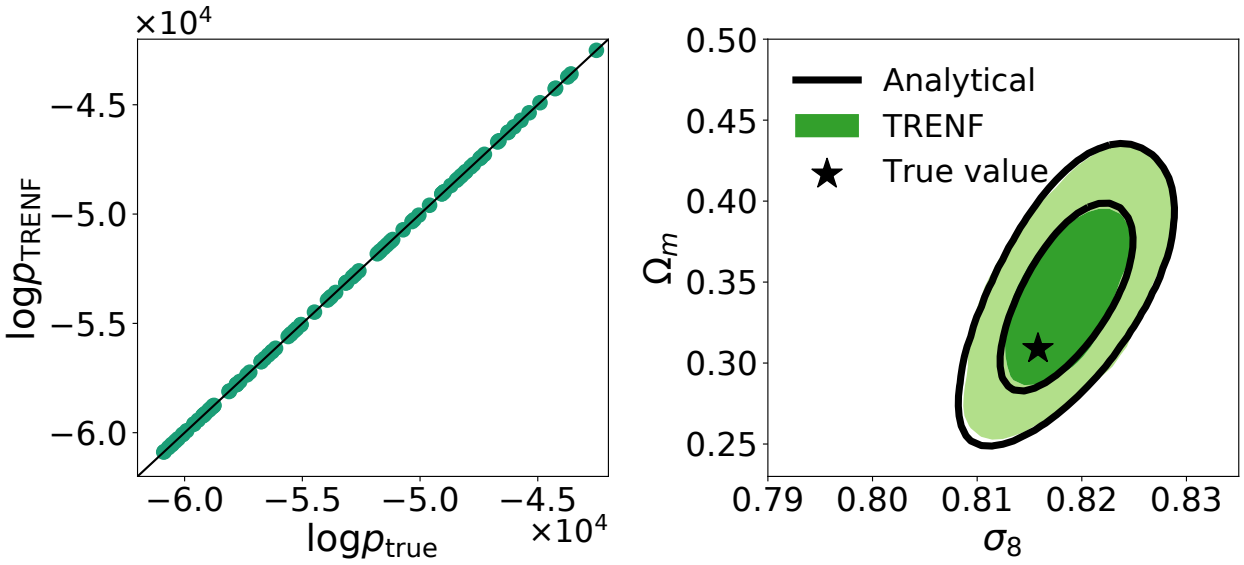



Figure 7.4: The comparison of log-likelihood (left panel) and posterior (right panel) between TRENF and analytical expression for Gaussian random fields from the test set.

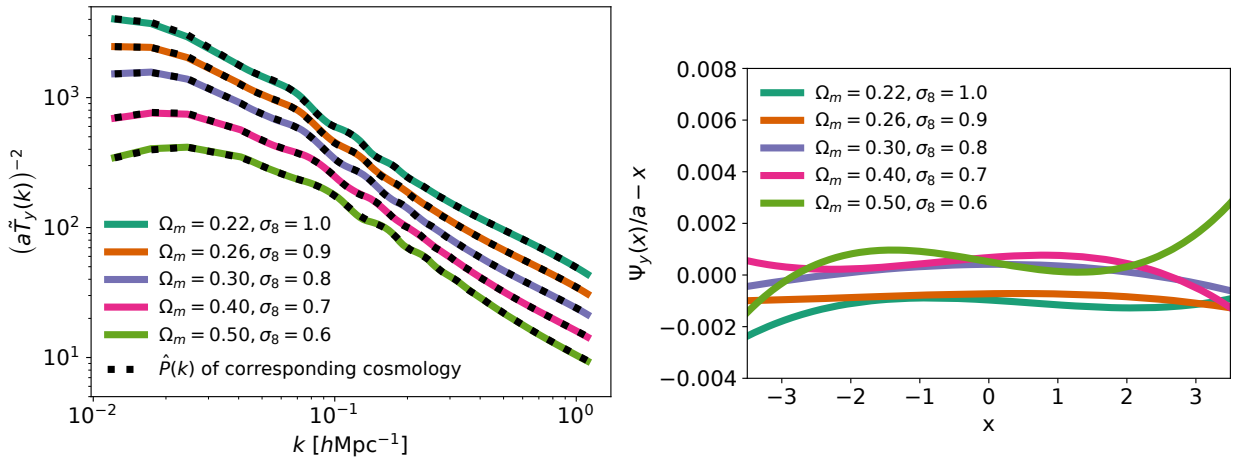



Figure 7.5: The comparison of the learned convolutional kernel $\tilde{T}(k)$ (top panel) and non-linearity $\Psi_y(x)$ (bottom panel) with the optimal solution (Equation 7.16 and 7.17) for GRFs.

$\sigma_8 \in [0.5, 1.1]$ [1]. We build a one-layer TRENF model with 16 spline points in $\tilde{T}$ and 8 spline points in $\Psi$. The architecture of the hyper network and the training process are the same as Section 7.3.

[1]For simplicity, we assume the numerical value of the 2D GRF power spectrum is equal to the 3D halofit matter power spectrum of the same $k$ amplitude, i.e., $P_{\rm GRF}(k) = P_{\rm halofit}(k) h{\rm Mpc}^{-1}$.

We compare the learned likelihood from TRENF with the true analytical likelihood

$$\log L_G = -\sum_k \frac{|\tilde{\delta}(k)|^2}{2\hat{P}_y(k)} - \frac{128^2}{2}\log(2\pi) - \frac{1}{2}\sum_k \log \hat{P}_y(k), \tag{7.15}$$

where $\hat{P}_y(k) = \frac{N^2}{L^2}P_y(k)$ is the covariance of the dimensionless $\tilde{\delta}(k)$, with $L = 512h^{-1}\mathrm{Mpc}$ denoting the box size and $N = 128$ denoting the mesh size. In Figure 7.4 we show the likelihood comparison on test data with random cosmology. We also show the posterior distribution from TRENF on test data and compare it with the posterior from the analytical likelihoods. The TRENF likelihood and posterior agree very well with the true answer, suggesting that TRENF is able to extract all the information from the GRFs.

In fact, the optimal solution of TRENF can be written down analytically for the GRFs:

$$\tilde{T}_y(k) = \frac{1}{a\sqrt{\hat{P}_y(k)}}, \tag{7.16}$$

$$\Psi_y(x) = ax, \tag{7.17}$$

where $a \neq 0$ is a free coefficient that represents the degeneracy between $\tilde{T}_y(k)$ and $\Psi_y(x)$, and which cancels out in Jacobian determinant. This allows us to explicitly check whether TRENF has found the optimal solution. In Figure 7.5 we show the learned $(a\tilde{T}_y(k))^{-2}$ and compare it with the scaled power spectrum $\hat{P}_y(k) = \frac{N^2}{L^2}P_y(k)$ for different cosmologies, where the coefficient $a$ is measured by fitting a linear relation between $x$ and $\Psi_y(x)$. We also present the difference between the learned $\Psi_y(x)$ and the true solution, $\Psi_y(x)/a - x$: the two agree with each other to about one part in a thousand across the entire range of $x$. These results demonstrate that the training of TRENF converges to the correct solution (Equation 7.16 and 7.17).

## Matter Overdensity Fields

We have shown that TRENF is able to learn the likelihood function accurately for GRFs. Now we explore the more challenging and more realistic dataset, the matter overdensity field. The dataset and the architecture of TRENF are the same as in Section 7.3. We optimize TRENF using the two-stage training strategy as described in Section 7.2 to improve the accuracy of posteriors.

In Figure 7.6 we present the $68\%$ and $95\%$ confidence regions of the posterior distribution on test data (we assume Planck 2015 cosmology parameters). We compare the posterior constraints from TRENF with the standard power spectrum analysis. TRENF models the full likelihood function of the data vector $x$ without any dimension reduction, so it provides much tighter constraints than the power spectrum, which only uses two-point function information. We measure the figure of merit, defined as the inverse of the area of $68\%$ confidence region, on 100 test data for both methods. We obtain 995 for TRENF and 176 for power spectrum, which means that TRENF significantly improves the posteriors relative to the power spectrum.

On 100 test data, there are 65 cases where the true cosmology is within the $68\%$ confidence region, and 95 cases the true cosmology is within the $95\%$ region. These numbers are consistent

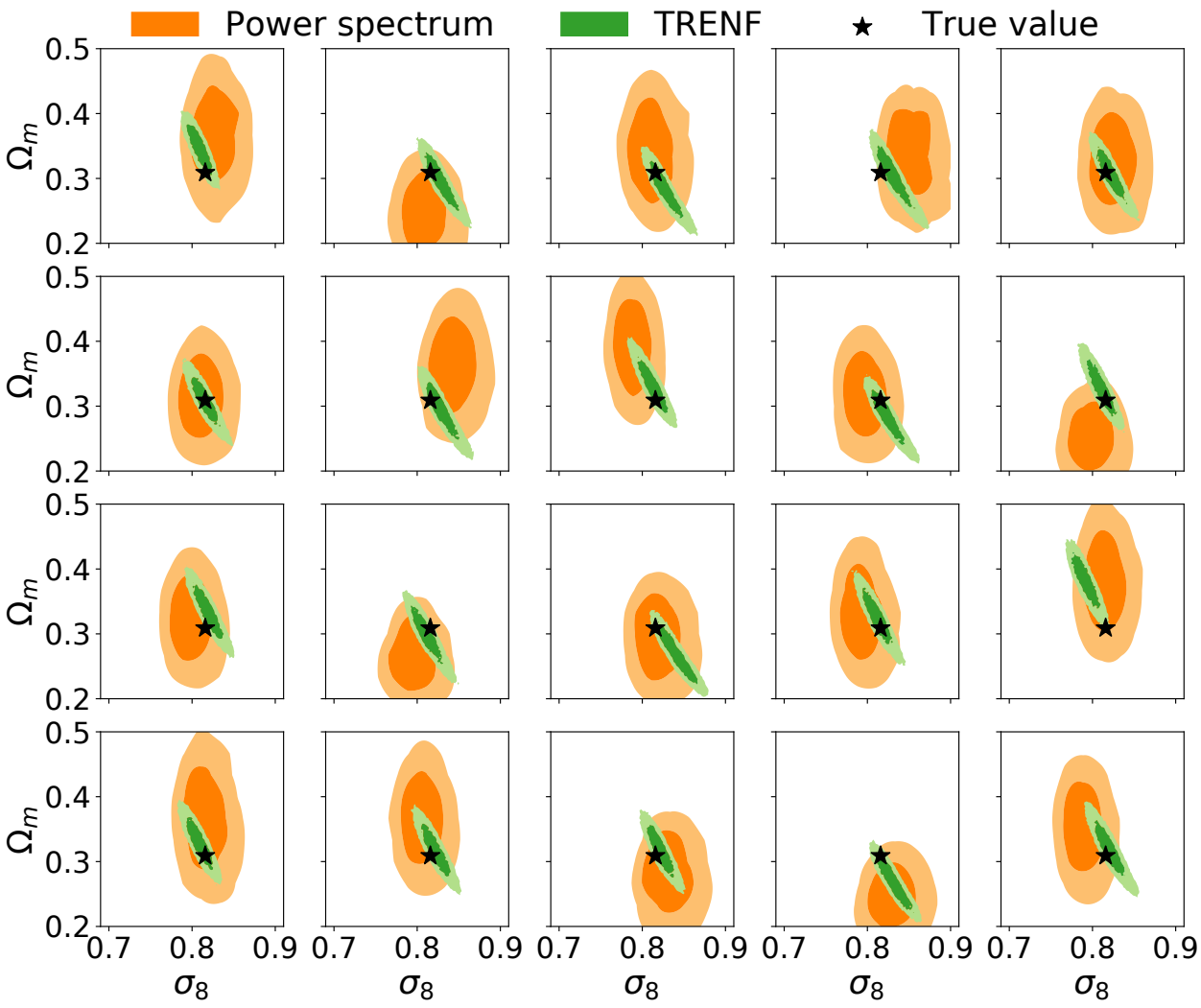



Figure 7.6: The posteriors from TRENF (green contour) and power spectrum (red contour) on uncurated test data. Figure of merit: 995 for TRENF, and 176 for power spectrum.

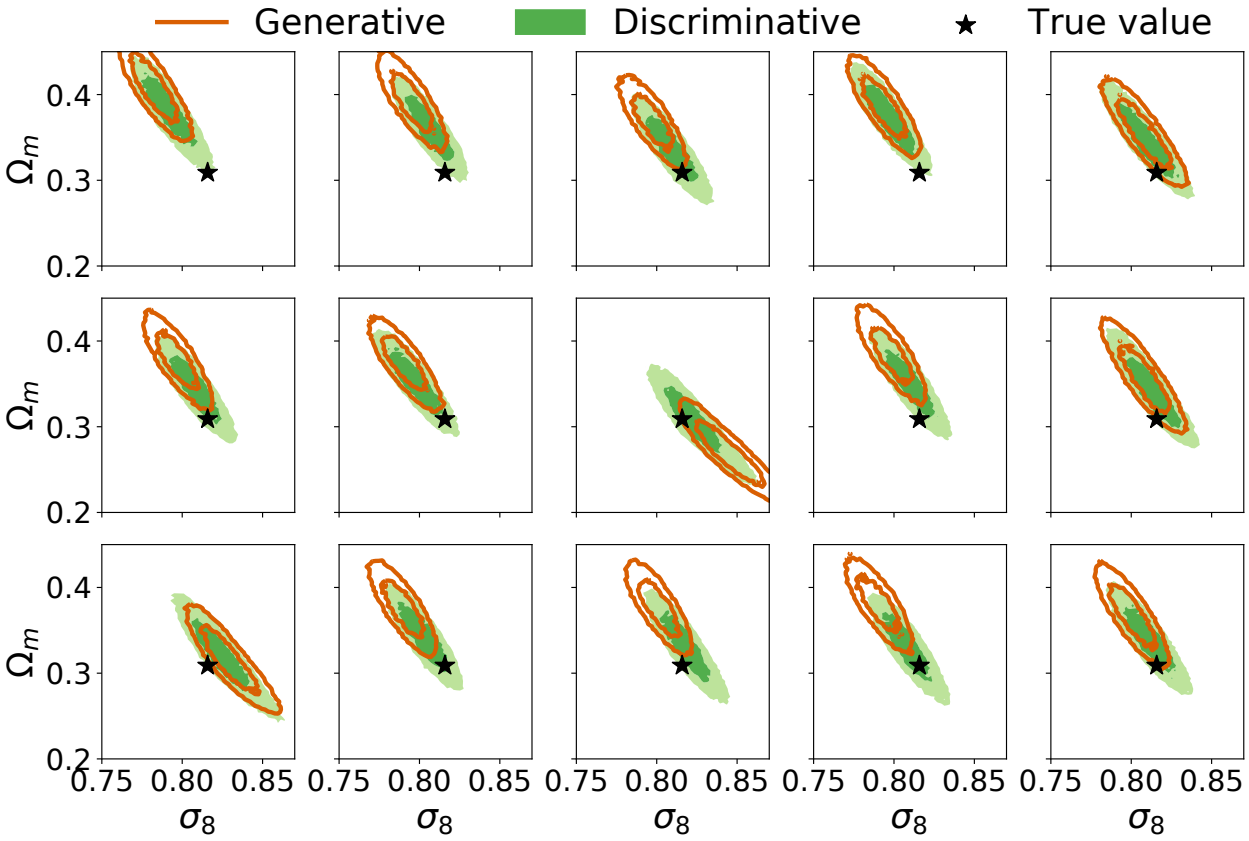



Figure 7.7: Comparison of the posterior constraints between generative learning (red contour) and discriminative learning (green contour) for the outlier cases in generative learning. Discriminative learning improves the posterior and results in properly calibrated posteriors, while generative learning is overconfident in the posteriors.

with the $68\%$ and $95\%$ expectation, suggesting that the posteriors from TRENF are properly calibrated. Note that we need second-stage discriminative learning to achieve this. If we train TRENF only with the generative loss, we find the model is overconfident and we get more than 5% outliers, where the true cosmology is outside the $95\%$ region, which is not consistent with the $5\%$ expectation. In Figure 7.7 we show some outlier cases from the generative learning, and we see that discriminative learning (two stage learning) improves the posterior.

## 7.5 Modeling the Data With Mask

In this section we present a generalization of TRENF to model the effects that break the data symmetry. As described in Section 7.2, our strategy is to first use an affine coupling layer to inpaint the missing pixels, and then introduce position dependence to the non-linearity $\Psi$ in Equation 7.6 to model these effects.

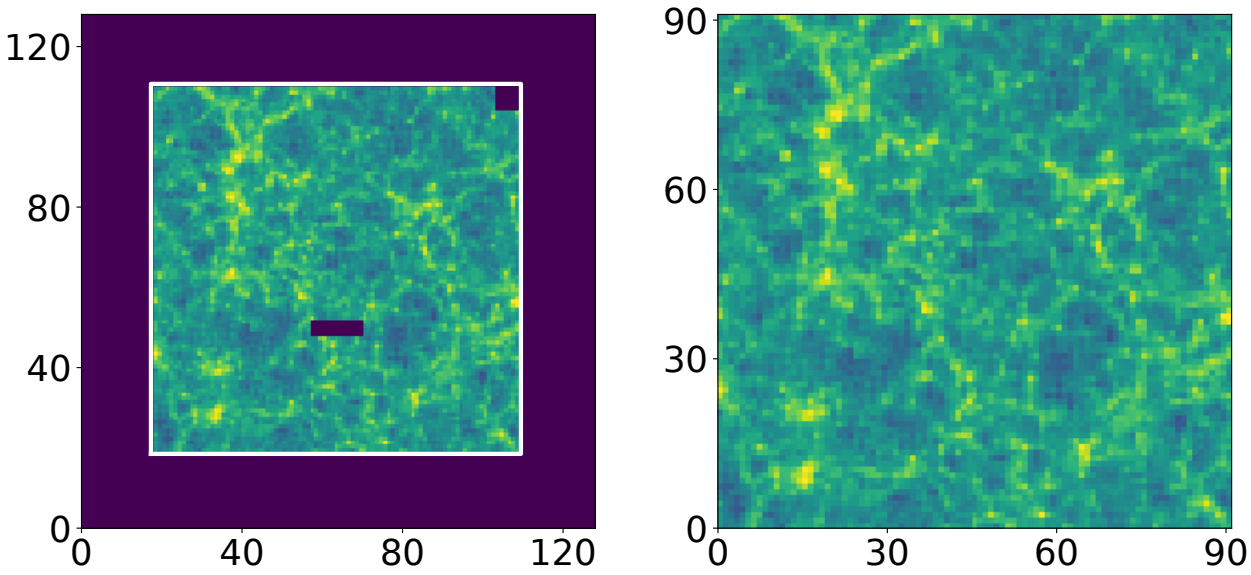



Figure 7.8: A visualization of the simplified mask we considered in this work (left panel), and the data after the affine coupling (inpainting) layer (right panel). In the left panel the blue regions represent missing pixels, and the white rectangle region denotes the data passed to the model.

**Dataset:** We consider a simple example where we take the central $91 \times 91$ pixels of the $128 \times 128$ matter overdensity fields so that the boundaries are no longer periodic, and then remove the $6 \times 6$ pixels at the upper right corner of the data to mimic the non-rectangular mask, and remove $4 \times 13$ pixels in the center of the data to mimic the observational effects, such as foregrounds, cosmic ray hits, detector failures, etc. We show a visualization of this mask in the left panel of Figure 7.8. During training and inference, we sample Gaussian noise in the $6 \times 6$ and $4 \times 13$ missing pixels, and the full $91 \times 91$ region is passed to the model for estimating the likelihood. Note that this is only a proof-of-principle study, and the mask we considered is a simplification to the realistic survey mask, but the methods we developed here should also apply to more realistic survey masks.

**Model:** Similar to Section 7.3, we first apply an inverse softplus transform and a normalization transform on the observed pixels to remove the $[0, \infty)$ boundary and rescale. We then apply affine coupling transforms (Equation 7.11) on the Gaussian noise of the two missing regions. After the inpainting, we add 5 layers of convolutions and position-dependent non-linearities. Here we have two hyper networks $g_{\tilde{T}}$ and $g_{\Psi}$ (Equation 7.12 and 7.13), and both of them, as well as the conditional networks in affine coupling layers, are chosen to be multilayer perceptrons with 2 hidden layers and $512$ neurons in each hidden layer. The other hyperparameters and training strategies are the same as Section 7.4.

We first show a visualization of the transformed data after inpainting in the right panel of Figure 7.8. Note that here we do not explicitly train the affine coupling layer to accurately recover the correct structures in the missing pixels. Instead, the goal of these layers is to inpaint structures

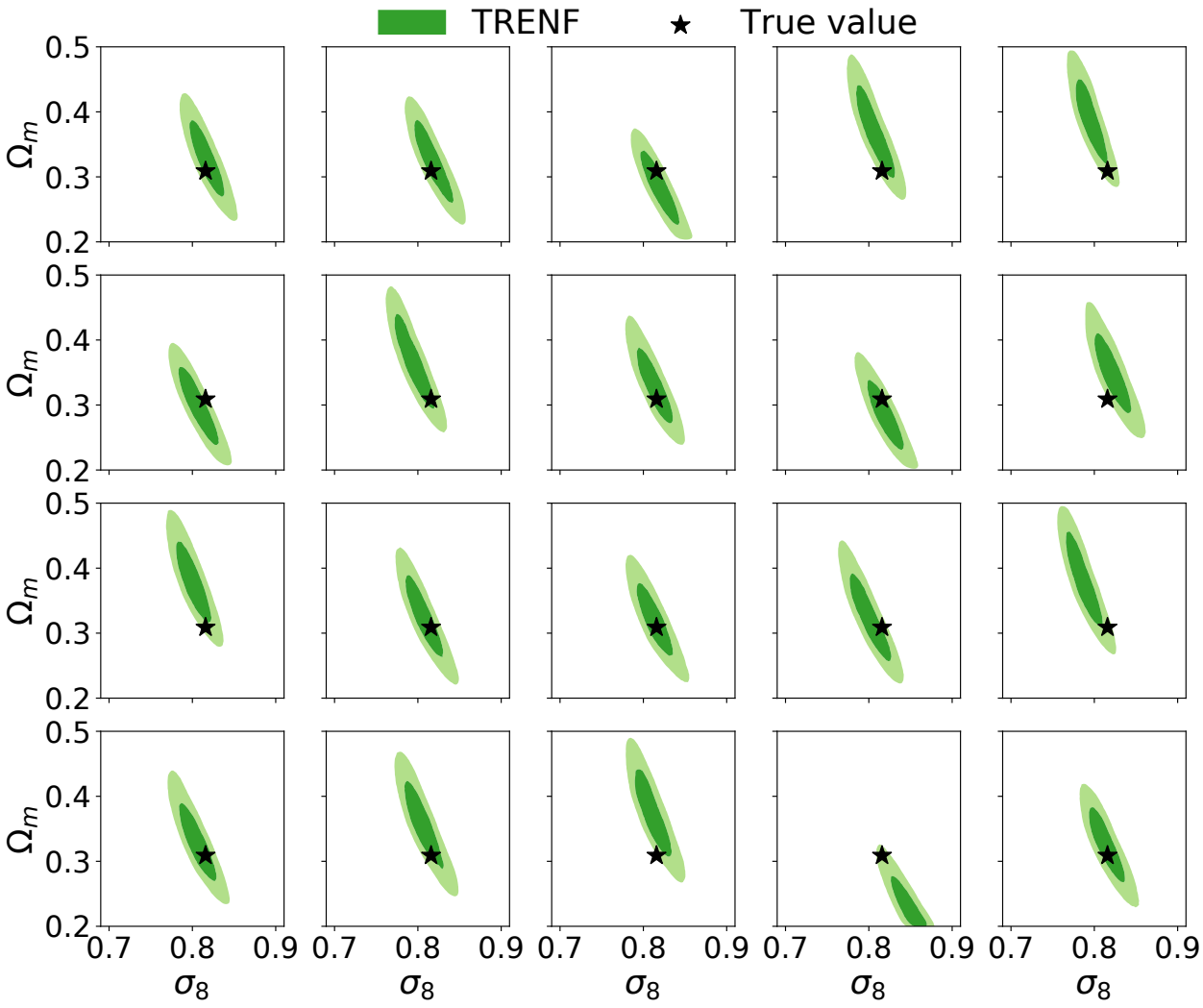



Figure 7.9: The posterior distribution from TRENF on test data with mask. $50\%$ of pixels are removed to mimic the survey mask. The figure of merit is 507.

that are statistically consistent with the observed data so that these missing pixels do not spoil the posterior analysis. To verify this, we show the posterior distribution of test data in Figure 7.9. We have verified that the uncertainty quantification is not miscalibrated: on 100 test data, there are 93 cases where the true cosmology is within the $95\%$ region, compared to the expected number of 95. If we assume that most of the information comes from the small scales, we expect that the amount of information would roughly be proportional to the area of the survey. Here we have removed about $50\%$ of the area as compared to the original dataset, so the constraining power should also be reduced by this amount. This is confirmed by our experiment: the figure of merit is now 507, as compared to 995 of the original dataset. This suggests that the amount of information extracted by TRENF is still close to optimal in the presence of the mask.

## 7.6 Beyond spherical kernels

As discussed at the end of Section 7.2, TRENF can be viewed as a CNN with the number of channels $c = 1$. In this section we will discuss TRENF in the framework of Steerable CNNs [48, 349, 47], which provides a general theory for equivariant networks. A steerable CNN defines the feature maps as steerable feature fields $v : \mathbb{R}^2 \to \mathbb{R}^c$. Under translation $t$ and rotation $r$, a steerable feature field $v(x)$ is transformed to $[\pi(tr)v](x)$, given by

$$[\pi(tr)v](x) = \rho(r) \cdot v(r^{-1}(x - t)), \tag{7.18}$$

where $\rho(r)$ is the type of the feature field and is a representation of the symmetry group. For example, in TRENF we have $c = 1$ and the feature field is a scalar field, which corresponds to the trivial representation $\rho(r) = 1$. In general steerable CNNs one can also have vector fields, where $\rho(r) = r$ is the standard representation and Equation 7.18 becomes the familiar transformation law of vector fields. We refer the readers to [48] and [47] for more details about steerable CNNs.

It has been shown that the most general linear map between steerable feature fields with type $\rho_{\text{in}}$ and $\rho_{\text{out}}$ is given by convolutions with kernel $T(x)$ satisfying [349, 47]

$$T(rx) = \rho_{\text{out}}(r)T(x)\rho_{\text{in}}(r^{-1}). \tag{7.19}$$

In normalizing flows the transformation is invertible, so the dimensionality of the feature fields should stay the same between different layers and the representation $\rho_{\text{in}}$ and $\rho_{\text{out}}$ must all be one dimensional. For $\mathrm{O}(2)$ group (rotation and reflection) the only one dimensional real representation is the trivial representation $\rho = 1$ [348], so we have $T(rx) = T(x)$ for any rotation $r$. Therefore spherical kernel is the only allowed kernel in TRENF.

There are several motivations to go beyond spherical kernels. One is to account for the effects of the grid, which breaks isotropy. Another is to approximate the $\mathrm{O}(2)$ symmetry with $\mathrm{D}_N$ group (discrete rotations by angles multiple of $\frac{2\pi}{N}$ and reflection). With a sufficiently large $N$, $\frac{2\pi}{N} \to 0$ and its multiples can approximate any angles. $\mathrm{D}_N$ group has several 1D real representations. We refer the readers to Appendix F.2 and Table 12 of [348] for irreducible representations of $\mathrm{D}_N$ group, as

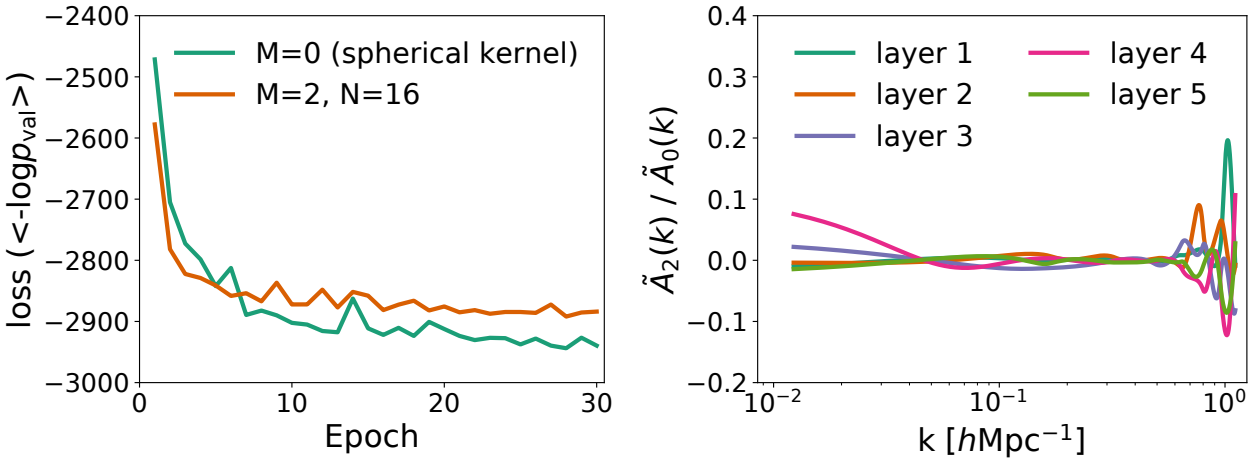



Figure 7.10: Left panel: the loss curve of TRENF with $M = 0$ (spherical kernel) and $M = 2, N = 16$ on validation set. Right panel: the amplitude ratio $\tilde{A}_2(k)$ / $\tilde{A}_0(k)$ of the convolution kernels in different layers of TRENF with $M = 2, N = 16$. See equation 7.21 for the definition of kernels we used in this experiment. TRENF with angular dependency $M = 2$ performs worse than TRENF with spherical kernels, and the amplitude of non-isotropic terms is small compared to the spherical symmetric term $\tilde{A}_0$.

well as all possible convolutional kernels between different $\mathrm{D}_N$ representations. For simplicity and invertibility considerations, we explore convolution kernels of the following forms:

$$T(r, \phi) = A_0(r) + \sum_{t=1}^{M} A_t(r) \cos(tN\phi), \tag{7.20}$$

where $A_0(r)$ and $A_t(r)$ are arbitrary radial functions. The above kernel ensures that the feature fields are all scalar fields and are equivariant under transformations of $\mathrm{D}_N$ group. In Fourier space, the kernel can be written as

$$\tilde{T}(k, \theta) = \tilde{A}_0(k) + \sum_{t=1}^{M} (-i)^{tN} \tilde{A}_t(k) \cos(tN\theta), \tag{7.21}$$

where $\theta$ is the polar angle of the Fourier $\boldsymbol{k}$ mode, and $\tilde{A}_t(k)$ is given by the Bessel function $J_{tN}$:

$$\tilde{A}_t(k) = \int J_{tN}(kr) A_t(r) r dr. \tag{7.22}$$

Similar to Section 7.2, we directly parametrize $\tilde{A}_t(k)$ in Fourier space using cubic splines and they will be learned from the data. We replace the spherical TRENF kernel with Equation 7.21 and apply the model on the 2D matter overdensity fields as described in Section 7.3. We tested different $N$ and $M$ choices ($N = \{8, 16, 32\}$, M=$\{$1,2,3$\}$), and compare their performance with the model with

spherical kernels. For TRENF with high order terms $M > 0$ we increase the width of the hyper network from 512 to 1024 for better conditional modeling, and keep the other hyperparameters the same. We find no improvements in terms of model loss (averaged negative data log-likelihood), sample quality, and posterior figure of merit. In Figure 7.10 we show that TRENF with spherical kernels converges to a better loss than TRENF with $M = 2, N = 16$. We also show that high order term $\tilde{A}_2$ is relatively unimportant compared to the spherical symmetric term $\tilde{A}_0$. On small scales $\tilde{A}_2/\tilde{A}_0$ deviates from 0, and this is probably because of imperfect optimization. $\tilde{A}_2/\tilde{A}_0 = 0$ is a better solution (left panel of Figure 7.10), but the model doesn't find it due to nonconvex optimization. This suggests that for the datasets we considered in this work, introducing extra angular dependence into the convolution kernel is not helpful and the spherical kernel is enough.

Another way to combine steerable CNNs with NFs is to adopt affine coupling transforms [71] and use steerable CNN for the coupling network. The coupling network does not need to be invertible and one can use more complex kernels with multidimensional steerable feature fields. However, affine coupling transforms require splitting the data into two parts, and such splitting is not generally equivariant. This approach is beyond the TRENF architecture, and we will leave it for future studies.

## 7.7 Discussion

The main goal of this paper is to develop a Normalizing Flow with built in translation and rotation symmetry (TRENF), for the purpose of generating new samples and data likelihood analysis in cosmology. We argue that cosmological fields are high dimensional distributions which cannot be approximated by low dimensional manifolds, so low-dimensional-manifold models such as GANs or VAEs are not appropriate for this application. In addition, NFs give direct access to the data likelihood, which is of crucial significance for the data analysis. Because NFs preserve the dimensionality of the data, they usually do not scale well to high dimensions compared to other generative models such as GANs or VAEs, but here we argue that taking advantage of translation and rotation symmetries makes this task significantly easier.

We argue that TRENF enables a clear path towards optimal cosmological analysis of the data, with a simple and computationally tractable approach. Specifically, we have shown that TRENF with only five layers saturates the information content that can be extracted from this architecture when applied to the 2D projections of cosmological N-body simulations, in the sense that adding more layers and using more complex kernels does not further improve the results. While proving optimality of the transport map from one distribution to another is a notoriously difficult problem for high dimensional distributions, several lines of argument suggest our approach enables near optimal analysis for the application developed in this paper.

First, when the method is applied to the Gaussian Random Fields with a known analytic solution, it extracts all the information correctly and optimally. Second, the inverse map from the data to the latent space is statistically indistinguishable from the target distribution of Gaussian white noise when we use correct cosmological parameters, while when the parameters are incorrect the map deviates from its target. This means that the non-Gaussian structures such as voids, filaments,

and halos have all been mapped into Gaussian white noise, and all of their information has been transferred via the Jacobian of the transformation into $p(x|y)$. TRENF extracts all the information, including from the two-point correlations, one point distribution, and bispectrum, which can be observed from the fact that these statistics become non-informative once the data are mapped into the latent space. The main power of TRENF is that it is able not only to extract information from so many different statistics, but do it in a way that can be optimal, and for which the output is the data likelihood itself.

TRENF can easily deal with noise in the data if noise is independent of position, as it preserves translation and rotation symmetry, and we can simply train TRENF on data with noise. A more difficult problem is that of cosmological data likelihood analysis in the presence of the survey mask. The mask breaks the translation and rotation symmetries, and is a notoriously difficult problem even for Gaussian fields: in the absence of the mask, the problem can be solved with Fast Fourier Transforms with $O(N \ln N)$ for flat geometry or $O(N^{3/2})$ for spherical geometry, while with the mask it requires a linear algebra solution that scales as $O(N^3)$, which becomes prohibitively expensive for large surveys. We introduce position dependence to the pointwise nonlinearity at each layer as the non-equivariant component in our model. We show that this approach is fast to train and to evaluate the likelihood. The constraining power (figure of merit) is reduced by the fraction of the area of the mask, consistent with our expectations if we assume most of the information comes from the small scales. This suggests that our approach is still optimal when the survey mask is included. Note that this method can also model other processes that break translation and rotation symmetry, for example, position-dependent noise, seeing, foregrounds, etc.

Potential applications and future generalizations of TRENF are numerous, here we list a few examples:

- TRENF enables the possibility of fast training and generation of new cosmological data outputs from a few existing simulations. This has numerous applications such as Lyman alpha forest, 21cm, and other intensity maps, weak lensing maps, projected galaxy clustering, X-ray and thermal SZ maps, etc. We expect that TRENF can learn efficiently with fewer training input maps (less data complexity), as compared to previous generative approaches, a consequence of translation and rotation symmetry built into the model. TRENF should also be generalized to output multiple maps of different tracers on the same area of the sky.

- TRENF training of latent space does not directly impose a spatial structure, although in practice we observe a strong correlation between the latent space and data space. It may be possible to make that more explicit, by enforcing the latent space to be the initial conditions of a simulation. In addition, one can also train TRENF as a function of time. In this case, TRENF would become an Eulerian N-body or hydro simulation. TRENF can also be used to learn the velocity field jointly with density, to describe the full phase space information.

- TRENF has the ability to perform nearly optimal posterior inference analysis of cosmological parameters via the data likelihood, given the TRENF's ability to evaluate $p(x|y)$. We emphasize that TRENF learns the data likelihood directly, and there is no need to learn the probability distribution of the summary statistics. One can view TRENF as a way to optimally

combine all of the summary statistics proposed in the cosmology literature, such as two-point, three-point, and higher order correlations, one-point distribution of various smoothing scales, void profile and void-void correlations, void and halo mass functions, topological statistics, etc. To the extent that the data have been mapped to Gaussian white noise, all of these summary statistics have been used optimally, by extracting their contribution to the Jacobian in $p(x|y)$.

- In addition to cosmological parameters, one can also train TRENF on astrophysical parameters, such as baryonic processes. These can be incorporated into the full likelihood analysis, by training TRENF on hydrodynamic simulations [336] or baryon maps generated by fast machine learning [54, 53], or semi-analytical approaches [11]. Once we have the likelihood of the data as a function of these parameters we can marginalize over these effects in the posterior analysis.

- TRENF can be generalized to 3D galaxy redshift space data, where observed redshift is the sum of position and velocity of the galaxy, and we no longer have exact rotation symmetry. Instead, we must describe the data in terms of the line of sight and perpendicular to the line of sight coordinates or their harmonic transforms, similar to our 2D expansion in equation 7.21.

- TRENF can be used to search for primordial non-Gaussianity in the latent space. Because the non-Gaussianity from the nonlinear evolution of structure is eliminated in the latent space, it becomes easier to search for other non-Gaussian effects, such as primordial non-Gaussianity.

If the promise of TRENF can be realized, its payoff will be the optimal extraction of cosmological information from the cosmology data. The remaining question, that applies to all cosmology LSS analyses, not just ours, is one of robustness: how do we know which information is reliable, and which is not, because it is corrupted by effects that are not included in the model? One way to address this is by separation of scales, with very small scale information likely being hopelessly corrupted by many astrophysical nuisance effects, and very large scale information likely being very robust. This argues for a scale-dependent version of TRENF, where we can verify robustness and optimality as a function of scale so that we can decide what scales to include and what to exclude. In the future, we plan to develop TRENF along these lines.

# Acknowledgements

We thank Patrick McDonald for helpful discussions. The material is based upon work supported by the National Science Foundation under Grant Numbers 1814370 and NSF 1839217, by NASA under Grant Number 80NSSC18K1274, and by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research under Contract No. DE-AC02-05CH11231 at Lawrence Berkeley National Laboratory to enable research for Data-intensive Machine Learning and Analysis.

# Chapter 8

# Multiscale Flow for Robust and Optimal Cosmological Analysis

The contents of this chapter were originally published in [55],

Multiscale Flow for Robust and Optimal Cosmological Analysis
*Dai B., Seljak U.*(arXiv:2306.04689) PNAS 121.9 (2024): e2309624121

We propose Multiscale Flow, a generative Normalizing Flow that creates samples and models the field-level likelihood of two-dimensional cosmological data such as weak lensing. Multiscale Flow uses hierarchical decomposition of cosmological fields via a wavelet basis, and then models different wavelet components separately as Normalizing Flows. The log-likelihood of the original cosmological field can be recovered by summing over the log-likelihood of each wavelet term. This decomposition allows us to separate the information from different scales and identify distribution shifts in the data such as unknown scale-dependent systematics. The resulting likelihood analysis can not only identify these types of systematics, but can also be made optimal, in the sense that the Multiscale Flow can learn the full likelihood at the field without any dimensionality reduction. We apply Multiscale Flow to weak lensing mock datasets for cosmological inference, and show that it significantly outperforms traditional summary statistics such as power spectrum and peak counts, as well as novel Machine Learning based summary statistics such as scattering transform and convolutional neural networks. We further show that Multiscale Flow is able to identify distribution shifts not in the training data such as baryonic effects. Finally, we demonstrate that Multiscale Flow can be used to generate realistic samples of weak lensing data.

## 8.1 Introduction

Numerous upcoming cosmological weak lensing surveys such as Euclid, the Vera Rubin Observatory (Rubin), or Nancy Grace Roman Space Telescope (Roman) hold the promise of revolutionizing our understanding of the universe, its origins, content, and its future evolution. How to efficiently extract the maximum amount of cosmological information from these data is a long-standing ques-

tion in large-scale structure (LSS) analysis. Due to the high-order correlations induced by nonlinear gravitational effects, the late-time cosmological fields are highly non-Gaussian with no tractable likelihood functions. Extracting information from these non-Gaussian fields has been mainly attempted through a limited set of summary statistics, with the most popular ones being the N-point correlation functions [e.g., 245, 246, 286, 290, 98]. However, while the two-point function is a natural choice even in the nonlinear regime, higher-order correlation functions are significantly more difficult to use due to the large number of statistical coefficients, large variance and high sensitivity to outliers [158]. Numerous other statistics have been proposed, including correlation functions on transformed or marked fields [234, 352], peak counts [144, 176], void statistics [353, 255], Minkowski functionals [220], scattering transform coefficients [44, 5], statistics learned by convolutional neural networks (CNNs) and other NNs [93, 41, 202, 112, 149], and many others. These analyses have the same underlying issues of summary statistics being ad-hoc and potentially sub-optimal. They require building effective likelihood functions from summary statistics using multi-variate Gaussian or Simulation-Based Inference (SBI) methods [51], which can be costly when the number of summaries is large. An alternative approach is using the reconstruction of initial conditions and estimating the field-level likelihood function by marginalizing over all possible initial conditions using a variety of methods such as sampling or optimization [147, 163, 346, 287, 259]. These methods are expensive because they perform reconstructions or sampling of 3-dimensional fields. They are also not well matched to the problem when the data is 2-dimensional, such as weak lensing.

Recently, Dai & Seljak [56] proposed directly learning the field-level data likelihood with Normalizing Flows (NFs). This approach does not require compressing the data into a low-dimensional summary statistic, and instead tries to extract all the information in the data from the field-level likelihood. Unlike the 3-d reconstructions, this approach does not require evaluating the high dimensional integral, and computes the likelihood function in a single forward pass of the flow network. Unlike SBI, it uses field level likelihood instead of summary statistics, performing Simulation Based Likelihood Inference (SBLI). To reduce the degrees of freedom when modeling the high-dimensional likelihood of the data they enforce translation and rotation symmetry into the NF. The resulting Translation and Rotation Equivariant Normalizing Flow (TRENF) agrees well with the analytical solution on Gaussian Random Fields, and it leads to significant improvement over the standard power spectrum analysis on nonlinear matter fields from N-body simulations [56]. Similarly, NFs with different architectures have been applied to neutral Hydrogen (HI) maps for fast sample generation and cosmological inference [121, 97].

Despite the differences in these LSS analysis methods, they all face the same challenge of robustness: how do we know which information is reliable, and which is not, if it is corrupted by effects that are ignored or inaccurately modeled? How do we detect distribution shifts in the actual data that were not in the training data? For example, most of these methods require accurate predictions from simulations, yet different hydrodynamical simulations and baryon models are not quite consistent with each other [83, 138]. Villaescusa-Navarro et al. [333] train CNNs to predict cosmological parameters from gas temperature maps. They find that their model, trained using IllustrisTNG simulations [253], fails dramatically when applied to gas maps produced by SIMBA simulations [61], due to the different subgrid models used in these two simulations. While

marginalizing over the baryon parameters, subgrid models and various systematic effects are helpful and necessary, there is no guarantee that current baryon and systematic models span all potential realistic scenarios.

One way to mitigate the impact of such modeling uncertainties is by separation of scales, with very small-scale information likely being contaminated by many astrophysical nuisance effects and observation systematics, and large-scale information likely being more robust. This strategy is widely used in current cosmological survey analyses of power spectrum or correlation function, for example by directly removing the small-scale information with scale cuts [e.g., 177, 313, 178], or by performing consistency checks between different scales [76]. The ability to perform a scale-dependent analysis is viewed as a distinct advantage of power spectrum or correlation function analysis when compared to other statistics.

In this paper, we apply the scale separation idea to the field-level likelihood modeling with NFs. Specifically, we use a set of scale-separated basis functions to represent the pixelized data, and decompose the data likelihood function into the contributions from different scales. Performing consistency checks between different scales enables us to decide what scale to include and what to exclude. While the Fourier basis is theoretically sound and widely used in such analysis, its kernels are not local in pixel space and require additional procedures in the presence of survey masks [56]. In this work we use a wavelet basis, which is localized in both real space and Fourier space, allowing us to easily handle the survey mask and to separate the signals from different physical scales. Such decomposition is also known as Multiresolution Analysis (MRA) in image processing. Furthermore, our hierarchical analysis also combines likelihood information from different scales to achieve optimality in the limit of sufficient training data.

## 8.2 Multiresolution Analysis with Fast Wavelet Transform

In this section we briefly introduce Multiresolution Analysis (MRA), which hierarchically decomposes the data into components at different scales, allowing us to separate the information from different scales and study them individually. This is particularly beneficial for cosmological analysis, since on large scales the universe can be modeled with simple physics and the data analysis is robust, while on small scales modeling the structure formation is harder due to nonlinear gravitational and astrophysical effects.

MRA is usually performed with Fast Wavelet Transform (FWT) [204]. While similar in concept to the Fourier basis, wavelet bases are constructed to be localized spatially, which is beneficial when analyzing maps with irregular footprints. Wavelet transform has been widely used in astronomical image processing [304] and statistical description of cosmological fields [44, 5]. In this work, we focus on decimated wavelet transform, which preserves the dimensionality of the data and can be viewed as a special kind of NF transforms.

The basic idea of FWT is to recursively apply low-pass filters (also called scaling functions) and high-pass filters (also called wavelet functions) to the data. In each iteration, the data $x_{2^n}$ with resolution $2^n$ is decomposed into a low-resolution approximation $x_{2^{n-1}}$, and detail coefficients of

the remaining signal $x^d_{2^{n-1}}$:

$$x_{2^{n-1}} = (\phi * x_{2^n}) \downarrow 2 \tag{8.1}$$
$$x^d_{2^{n-1}} = (\psi * x_{2^n}) \downarrow 2 \tag{8.2}$$

where $\phi$ is the low pass filter (scaling function), $\psi$ is the high pass filter (wavelet function), $*$ is the convolution operation, and $\downarrow 2$ is the operator to downsample the data by a factor of 2: $(x \downarrow 2)_{i,j} = x_{2i,2j}$. This is equivalent to a convolution with stride 2. For a 2D map $x_{2^n}$, we have three high pass filters to match the dimensionality, and the dimension of $x^d_{2^{n-1}}$ is $3 \times 2^{n-1} \times 2^{n-1}$. Then the low-resolution data $x_{2^{n-1}}$ is passed to the next iteration and treated as the input for further decomposition. Note that this decomposition is bijective and in each iteration the input data can be reconstructed with the inverse wavelet transform.

In this work, we use Haar wavelet [113], the simplest and the most spatially localized wavelet function. Its scaling function and wavelet function can be represented by the following $2 \times 2$ kernel in real space:

$$\phi = \frac{1}{4}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \psi_1 = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}, \tag{8.3}$$
$$\psi_2 = \frac{1}{2}\begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix}, \psi_3 = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}, \tag{8.4}$$

where we have scaled the scaling function such that $x_{2^{n-1}}$ is exactly the low-resolution version of $x_{2^n}$ by taking the average of every $2 \times 2$ patch. The localized kernel of the Haar wavelet allows us to handle the survey mask easily, but our method can be generalized to other more complicated wavelet transforms, e.g., Daubechies wavelets [59].

With MRA, the log-likelihood of a map $x_{2^n}$ with resolution $2^n$ can be rewritten into an auto-regressive form as

$$\begin{aligned} \log p(x_{2^n}|y) &= \log p(x_{2^{n-1}}, x^d_{2^{n-1}}|y) \\ &= \log p(x_{2^{n-1}}|y) + \log p(x^d_{2^{n-1}}|x_{2^{n-1}}, y) \\ &= \log p(x_{2^{n-2}}|y) + \log p(x^d_{2^{n-2}}|x_{2^{n-2}}, y) \\ &\quad + \log p(x^d_{2^{n-1}}|x_{2^{n-1}}, y) \\ &= \cdots \\ &= \log p(x_{2^k}|y) + \textstyle\sum_{m=k}^{n} \log p(x^d_{2^m}|x_{2^m}, y), \end{aligned} \tag{8.5}$$

where $2^k$ is the scale where we stops the decomposition, and $k$ can be any integer between 0 and $n$. In practice, we can choose $k$ such that it corresponds to the scale that either has extracted all the information from the data, or is large enough not to be affected by unknown small-scale systematic effects.

## 8.3 Multiscale Flow

### Normalizing Flows

Flow-based models provide a powerful framework for density estimation [71, 239] and sampling [161]. These models map the data $x$ to latent variables $z$ through a sequence of invertible transformations $f = f_1 \circ f_2 \circ ... \circ f_n$, such that $z = f(x)$ and $z$ is mapped to a base distribution $\pi(z)$. The base distribution $\pi(z)$ is normally chosen to be a Gaussian with zero mean and unit variance, $\pi(z) = \mathcal{N}(0, I)$. The probability density of data $x$ can be evaluated using the change of variables formula:

$$\begin{aligned} p(x) = & \quad \pi(f(x)) \left| \det \left( \frac{\partial f(x)}{\partial x} \right) \right| \\ = & \quad \pi(f(x)) \prod_{l=1}^{n} \left| \det \left( \frac{\partial f_l(x)}{\partial x} \right) \right| . \end{aligned} \tag{8.6}$$

To sample from $p(x)$, one first samples latent variable $z$ from $\pi(z)$, and then transform variable $z$ to $x$ through $x = f^{-1}(z)$. The transformation $f$ is usually parametrized with neural networks $f_\phi$, and the parameters $\phi$ are normally estimated using Maximum Likelihood Estimation (MLE):

$$\phi^* = \arg\max_{\phi} \frac{1}{N} \sum_{i=1}^{N} \log p_\phi(x_i), \tag{8.7}$$

where the data likelihood $p(x)$ is given by Equation 8.6. The MLE solution minimizes the Kullback-Leibler (KL) divergence between the model distribution $p_\phi(x)$ and the true data distribution. The parameterization of $f$ must satisfy the requirements that the Jacobian determinant $\det(\frac{\partial f_l(x)}{\partial x})$ is easy to compute for evaluating the density, and the transformation $f_l$ is easy to invert for efficient sampling.

In cosmological analysis we are interested in the likelihood function $p(x|y)$, which can be estimated using conditional Normalizing Flows (NFs). In conditional NFs the flow transformation is dependent on the conditional parameters $y$, i.e., $f = f_{\phi,y}$. We discuss below how we parametrize and train the conditional flow $f_{\phi,y}$.

### Multiscale Flow

With the likelihood decomposition Equation 8.5, our task now is to build NFs to model different likelihood terms separately. For simplicity, we will drop the subscript $2^m$ in this section, and simply refer to the conditional likelihood term $\log p(x^d_{2^m}|x_{2^m}, y)$ as $\log p(x^d|x, y)$. The model described here is similar to Wavelet Flow [359], even though they are developed independently. Following Glow [161], our flow transformation $f(x|y)$ consists of multiple block flows, where each block consists of an actnorm, an invertible $1 \times 1$ convolution, and an affine coupling layer (Fig. 8.1).

**Actnorm**: The actnorm layer applies an affine transformation per channel, similar to batch normalization [140], but its scale and bias parameters are initialized such that the output has zero mean and unit variance per channel given an initial minibatch of data, and then these parameters are treated

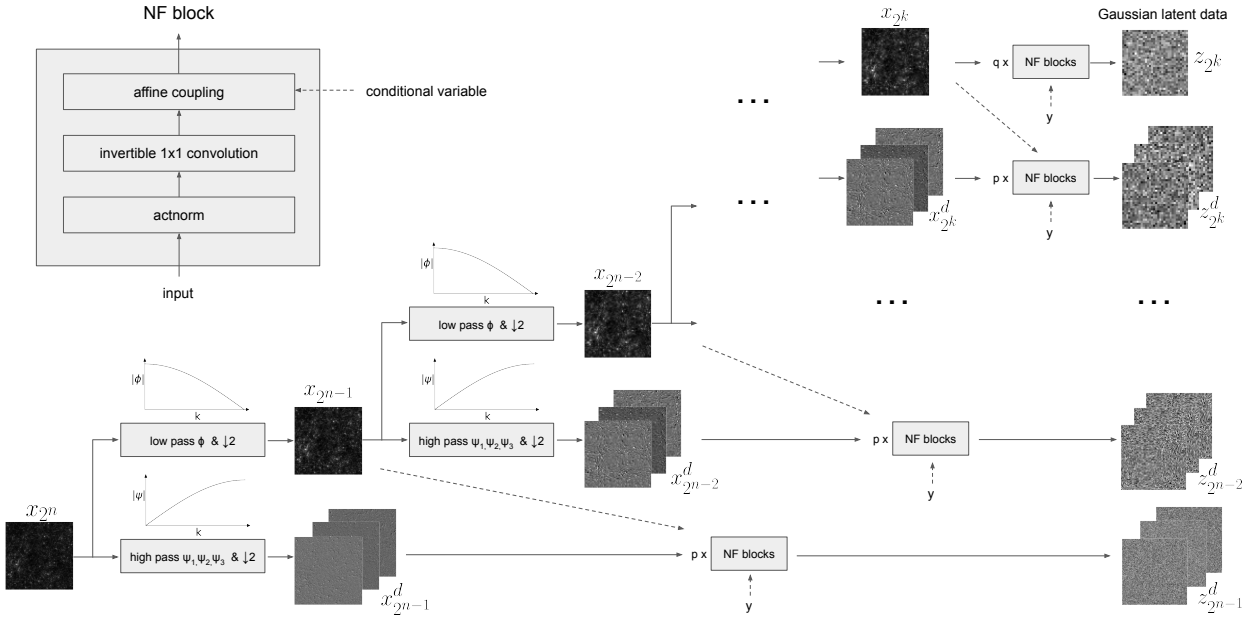



Figure 8.1: Illustration of Multiscale Flow model. The input map $x_{2^n}$ with resolution $2^n$ is iteratively processed with a set of low pass filters ($\phi$), high pass filters ($\psi_1$, $\psi_2$, $\psi_3$) and downsampling ($\downarrow 2$), resulting in a series of detailed maps $x^d_{2^{n-1}}, x^d_{2^{n-2}}, \cdots, x^d_{2^k}$ and an approximation map $x_{2^k}$. These maps are then transformed by several NF blocks to Gaussian latent maps $z^d_{2^{n-1}}, z^d_{2^{n-2}}, \cdots, z^d_{2^k}, z_{2^k}$, where each NF block is composed of an actnorm layer, an invertible $1 \times 1$ convolution, and an affine coupling layer (Equation 8.8, 8.9), as shown on the top left of this figure. The NF transformation is conditioned on the conditional variable $y$ and approximation maps, which are represented by dashed arrows in the illustration. The log-likelihood of the input map $x_{2^n}$ can be calculated with Equation 8.5.

as regular trainable parameters.

**Invertible** $1 \times 1$ **convolution**: The invertible $1 \times 1$ convolution is a learnable $C \times C$ matrix (where $C$ is the number of channels) that linearly mixes different channels.

**Affine coupling**: The affine coupling layer firstly splits the data $x^d$ to $x^{d1}$ and $x^{d2}$ based on the channels, and then applies pixel-wise affine transformation to $x^{d2}$, with scale and bias given by $x^{d1}$:

$$(\log s,\ t) = \quad \mathrm{CNN}(x^{d1}, x, y) \tag{8.8}$$

$$z^{d2} = \quad \exp(\log s) \cdot x^{d2} + t, \tag{8.9}$$

where $\log s$ and $t$ are scale and bias coefficient maps with the same dimensionality as $x^{d2}$, and CNN is a learned function parametrized by a convolutional neural network. The dependence of conditional parameter $y$ is modeled by introducing gating into CNN, i.e., each channel of CNN is scaled by a value between 0 and 1 which is determined by parameter $y$. This gating allows the conditional variable $y$ to determine the relative weights between different features (channels). The

output of the affine coupling layer is the concatenation of $x^{d1}$ and $z^{d2}$. In other words, the affine coupling layer applies an affine transformation to $x^{d2}$ and leaves $x^{d1}$ unchanged. In this paper, we consider 2D maps, so at each scale $x^d$ contains 3 maps (channels). We set the first channel to be $x^{d1}$, and the other two channels to be $x^{d2}$.

To summarize, a Multiscale Flow consists of multiple NFs, and each NF models one term of the likelihood decomposition (Equation 8.5) separately. The large-scale term $\log p(x_{2^k}|y)$ is modeled by $q$ flow blocks, and each other term $\log p(x^d_{2^m}|x_{2^m}, y)$ is modeled with $p$ flow blocks, where $p$ and $q$ are hyperparameters in the model. Note that all of these NFs can be trained independently in parallel to speed up the training process.

## Training

Following Dai & Seljak [56], we adopt a two-stage training strategy in this work: we first train the NF with the generative loss, which minimizes the negative log-likelihood and is the standard loss function of NF (Equation 8.7 with conditional variable y):

$$\mathcal{L}_{\rm g} = -\frac{1}{N}\sum_{i=1}^{N}\log p(x_i|y_i). \tag{8.10}$$

The generative loss is suitable for sampling and density estimation, but may lead to a biased or overconfident posterior[56]. To solve this issue they propose further optimizing the posteriors by training the model with the discriminative loss,

$$\begin{aligned}\mathcal{L}_{\rm d} &= -\frac{1}{N}\sum_{i=1}^{N}\log p(y_i|x_i)\\ &= -\frac{1}{N}\sum_{i=1}^{N}\left[\log p(x_i|y_i) + \log p(y_i) - \log p(x_i)\right],\end{aligned} \tag{8.11}$$

where the evidence $p(x)$ is estimated using Importance Sampling (IS):
$\log p(x) \approx \log \frac{1}{M}\sum_{y_j\sim q(y|x)}^{M}\frac{p(x|y_j)p(y_j)}{q(y_j|x)}$, and $q(y|x)$ is chosen to be a Gaussian distribution with learned mean and fixed covariance matrix. However, we find that IS becomes inefficient when the number of parameters $y$ gets large and when the posterior becomes non-Gaussian. In this work, we notice that

$$\begin{aligned}\nabla_\phi\mathcal{L}_{\rm d} = -\frac{1}{N}\sum_{i=1}^{N}\Bigg(&\nabla_\phi\log p_\phi(x_i|y_i) - \\ &\mathop{\mathbb{E}}_{y_j\sim p_\phi(y|x_i)}\nabla_\phi\log p_\phi(x_i|y_j)\Bigg),\end{aligned} \tag{8.12}$$

where we have used a trick that is commonly seen in the training of energy-based models. Its derivation can be found in [296]. In the training, we replace the expectation with a single Monte

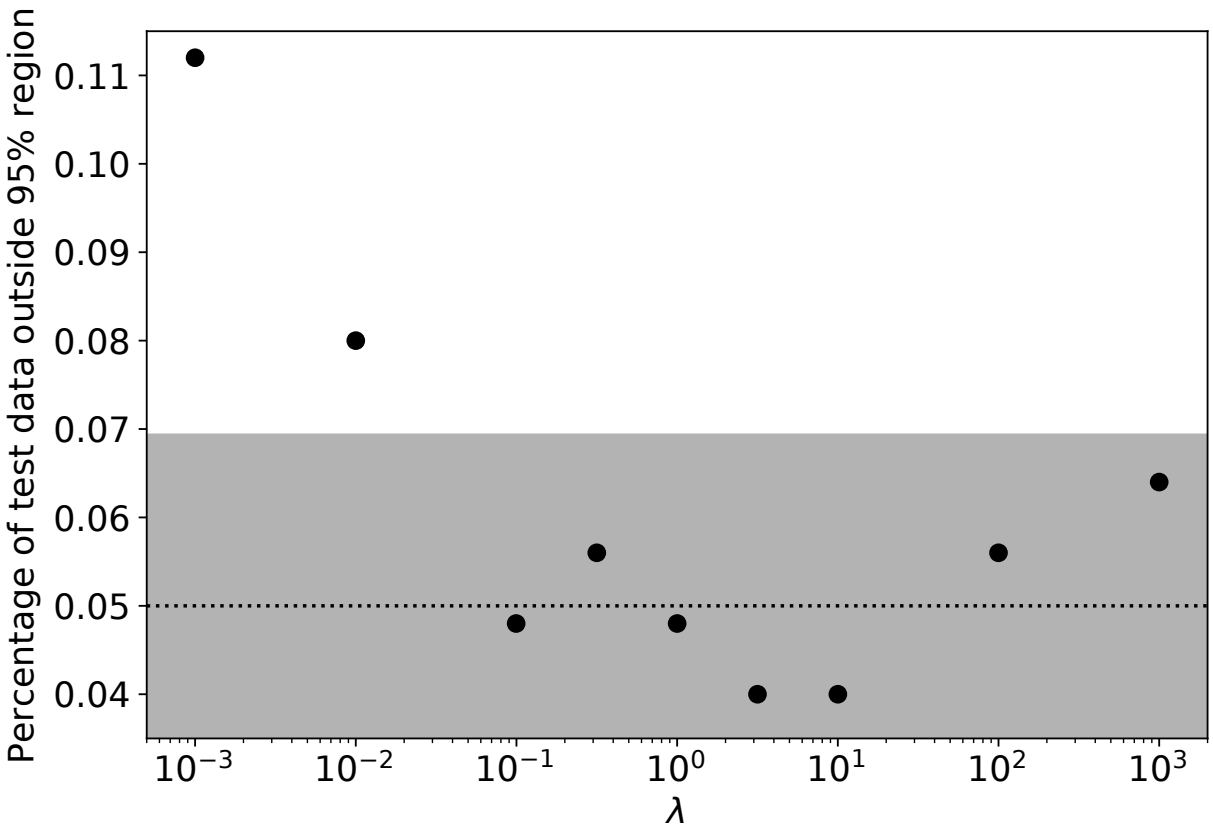



Figure 8.2: Percentage of test data that fall outside $95\%$ confidence region for different $\lambda$ values. A perfectly calibrated posterior has $5\%$ outliers (data that fall outside $95\%$ confidence region). The shaded region shows the measurement uncertainty due to the finite test data. This measurement is made on weak lensing maps with $64^2$ resolution and $n_g = 30\mathrm{arcmin}^{-2}$ galaxy density.

Carlo sample of the posterior $p(y|x_i)$, and we obtain these samples by running a Hamiltonian Monte Carlo (HMC) sampler [77]. These samples are saved, and then updated with a few HMC steps every epoch of training [319]. An advantage of this gradient formula compared to naively evaluating Eq 8.11 is that instead of evaluating the evidence term $\log p(x) = \log \int p(x|y)p(y)dy$, we now evaluate $\int \log p(x|y)p(y|x)dy$. The estimation of the former usually comes with a large variance, while the latter can be estimated with only a few HMC samples.

After the generative training, we add this loss to the generative loss with a hyperparameter $\lambda$,

$$\mathcal{L} = \frac{1}{1+w\lambda}\mathcal{L}_{\mathrm{g}} + \frac{w\lambda}{1+w\lambda}\mathcal{L}_{\mathrm{d}}, \tag{8.13}$$

where $w = \frac{d_x}{d_y}$ is a prefactor to balance the dimension difference between the data and the parameter space, and we divide the loss by $1 + w\lambda$ to normalize the weights. In Figure 8.2 we show the percentage of outliers in our posterior analysis with different $\lambda$ values. For very small $\lambda$ the posterior

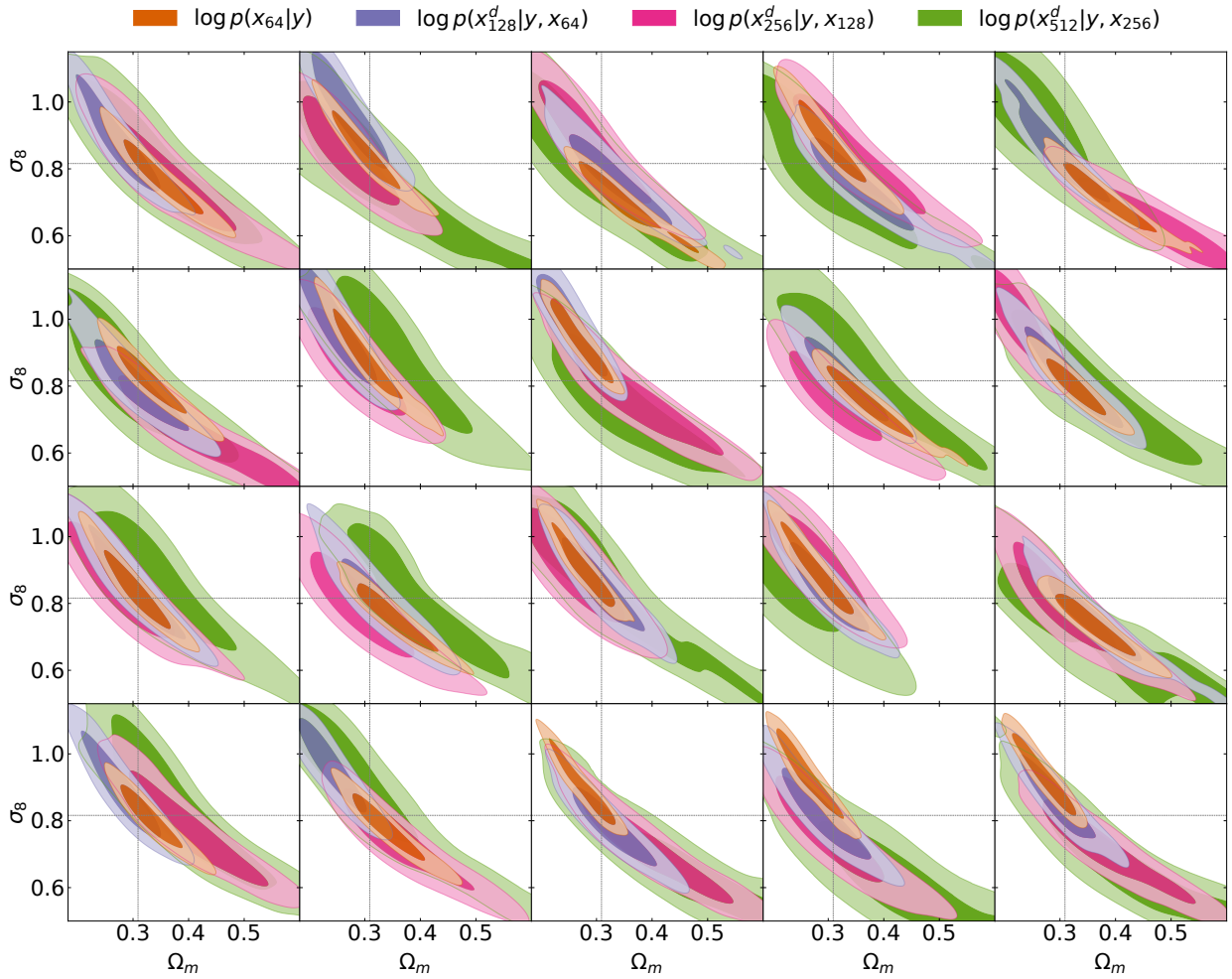



Figure 8.3: Multiscale Flow posterior comparison of different scales on 20 test data with galaxy number density $n_g = 30\text{arcmin}^{-2}$.

is too narrow (underestimated errors) and the loss is dominated by the first loss term (generative loss). For $\lambda > 0.1$ the posterior is well calibrated due to the second term $\tilde{\mathcal{L}}_\text{d}$. In this paper, we use $\lambda = 1$ to calibrate the posterior.

## 8.4 Results

### Cosmological constraints from noisy weak lensing maps

We apply Multiscale Flow to $3.5 \times 3.5\text{deg}^2$ mock weak lensing convergence maps [270] for field-level inference. We decompose the $512^2$ resolution map to four scales, with likelihood

Table 8.1: Comparison of the constraining power between different methods.

| $Method$ | $n_g = 10\text{arcmin}^{-2}$ | $n_g = 30\text{arcmin}^{-2}$ | $n_g = 100\text{arcmin}^{-2}$ |
|---|---|---|---|
| Multiscale Flow $p(x_{512}\|y)$ | **89** | **248** | **740** |
| Multiscale Flow $p(x_{256}\|y)$ | 82 | 226 | 631 |
| Multiscale Flow $p(x_{128}\|y)$ | 76 | 191 | 472 |
| Multiscale Flow $p(x_{64}\|y)$ | 62 | 130 | 298 |
| power spectrum | 30 (30) | 52 (51) | 81 (79) |
| peak count | (40) | (85) | (137) |
| CNN | (44) | (121) | (292) |
| scattering transform$s_0 + s_1 + s_2$ | ($\lesssim 50$) | ($\lesssim 140$) | ($\lesssim 329$) |

1. Unless specified with Multiscale Flow, the analysis of other approaches are performed on maps with resolution $512^2$.

2. The numbers in parenthesis are estimated using maps with $1$ arcmin Gaussian smoothing. We expect this smoothing to have little effect on constraining power estimation, because the small-scale modes are dominated by shape noise. This is also explicitly verified in the case of power spectrum, where we show FoM with and without smoothing. We have also verified that CNN produces comparable results with and without smoothing.

3. The FoM of the scattering transform is estimated using the Fisher matrix, which is an upper limit of the true FoM according to the Cramér-Rao inequality. It has been shown that Fisher forecast could potentially overestimate the 1D parameter constraints by a factor of 2, due to the non-Gaussian distribution of the statistics. [240].

decomposition

$$\begin{aligned}\log p(x_{512}|y) = \quad & \log p(x_{64}|y) + \log p(x^d_{64}|x_{64}, y) + \\ & \log p(x^d_{128}|x_{128}, y) + \log p(x^d_{256}|x_{256}, y).\end{aligned} \tag{8.14}$$

The posterior comparison of different scales on $20$ test maps with galaxy number density $n_g = 30\text{arcmin}^{-2}$ is shown in Figure 8.3. The posterior constraints of all scales are consistent with the true cosmological parameters, which are shown as black lines. The constraining power of Multiscale Flow of different galaxy shape noise levels is shown in Table 8.1. We list the figure of merit (defined as the reciprocal of the $1\sigma$ confidence area on the $(\Omega_m, \sigma_8)$ plane) of maps with different resolutions, and compare them with summary statistics power spectrum, peak count, scattering transform [44], and statistics learned by CNNs [270]. Multiscale Flow achieves the best performance among all methods, outperforming power spectrum by factors of 3, 5 and 9 on galaxy densities $n_g = 10, 30, 100\text{arcmin}^{-2}$, respectively. Multiscale Flow also achieves two to three higher constraining power when compared to peak counts, CNN, and scattering transform.

## Impact of baryons

Next, we apply Multiscale Flow to mock weak lensing maps with baryonic physics included [198]. Similar to the previous experiment, these maps also have a resolution of $512^2$, and we adopt the

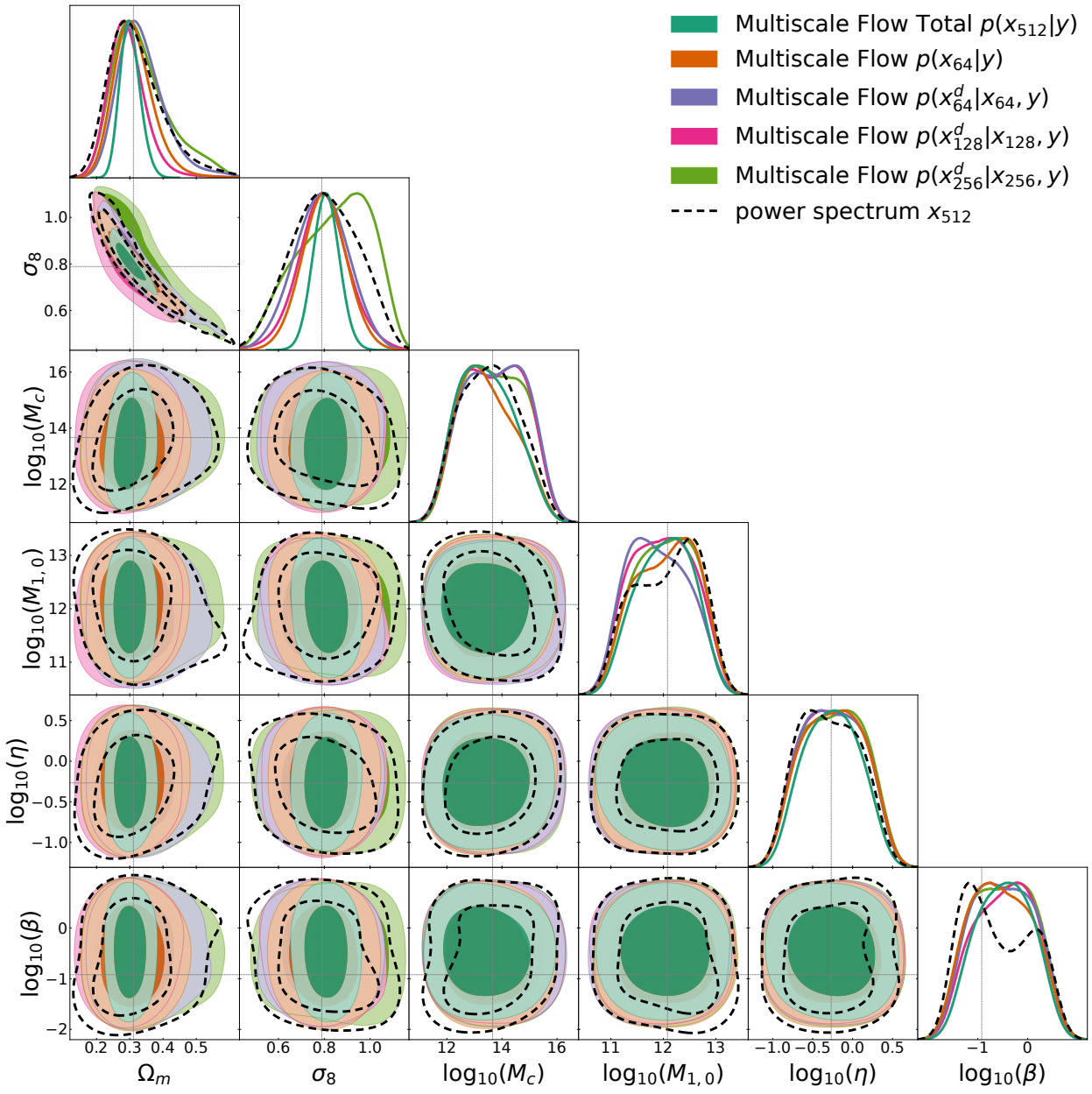



Figure 8.4: Comparison of posterior distributions between different scales of Multiscale Flow and power spectrum on a $3.5 \times 3.5\text{deg}^2$ convergence map with $n_g = 20\text{arcmin}^{-2}$.

same likelihood decomposition as Equation 8.14. We have 6 physical parameters in total, i.e., cosmological parameters $\Omega_m$ and $\sigma_8$, and 4 baryon parameters [11]. The posterior distributions of Multiscale Flow and power spectrum of a test data with $n_g = 20\text{arcmin}^{-2}$ are shown in Figure 8.4. In Table 8.2 we compare the constraining power of Multiscale Flow, power spectrum, and CNN

[198] on $(\Omega_m, \sigma_8)$ plane. With the presence of baryon physics, Multiscale Flow has $2.5-4$ times higher constraining power on cosmological parameters when compared to the power spectrum. It also outperforms CNN by a factor of 2.

Unfortunately, due to the small area of the lensing map, all these methods cannot constrain baryon parameters very well (see also Figure 5 of Lu et al. [198] for CNN constraints), and the posterior is dominated by the prior bounds, especially in the cases of high shape noise. Therefore, marginalizing the baryon parameters has a small impact on the Figure of Merit. With smaller shape noise and a more powerful model, the posterior becomes more dominated by likelihood rather than the prior, and the degradation of FoM when marginalizing over the baryon parameters gets larger. This explains why the degradation of baryon marginalization is larger for Multiscale Flow compared to the power spectrum, and why the degradation is larger in high galaxy number density cases. However, it is important to recognize that with better statistical power, and simpler baryonic models, we expect field level inference to be able to break the degeneracies between baryonic and cosmological parameters.

We apply Multiscale Flow to test data with fiducial parameters, and in Table 8.3 we report the percentage of test data with true cosmological parameters to fall in $68\%$ and $95\%$ confidence regions. In most cases the percentages are larger than the $68\%$ and $95\%$ expectation, suggesting that our posterior constraint is conservative.

## Identifying distribution shifts

Identifying distribution shifts from unknown effects that are present in the data, but not in the training simulations, is one of the great challenges of modern Machine Learning. Here we propose two different methods to identify such shifts. In the first approach, we evaluate the likelihood value of test data at MAP $\log p(x|y_{\rm MAP})$ and compare it with the distribution of training data. If it is smaller than the typical likelihood values of training data, it is likely not in the typical set of training distribution. In the second approach, we use consistency of information as a function of scale to identify such shifts. Specifically, we evaluate

$$\Delta \log p(x_m|y) = \log p(x_m|y_{\rm MAP}) - \log p(x_m|y_{{\rm MAP},m}), \tag{8.15}$$

where $x_m$ is the data of a specific scale, $y_{{\rm MAP},m} = \arg\max_y \log p(y|x_m)$ is the MAP of this scale, and $y_{\rm MAP}$ is the MAP of all the scales. If there are scale-dependent systematic effects that bias the posterior in different ways, we expect $y_{\rm MAP}$ and $y_{{\rm MAP},m}$ to be quite different, and $\Delta \log p(x_m|y)$ should be smaller compared to those of training data.

As a simple example, we train the Multiscale Flow with dark-matter-only convergence maps [270], and apply the model to convergence maps with baryon physics included [198]. We show the posterior distributions from different scales in the upper left panel of Figure 8.5. The baryon physics modifies the matter distribution on small scales and biases the posterior constraints from small scales. In this case, naively combining all of the scales leads to a posterior constraint that is $2\sigma$ biased (dark green contour). The inconsistency of posterior between different scales suggests a presence of unknown systematics (baryon physics) that is not modeled in the training

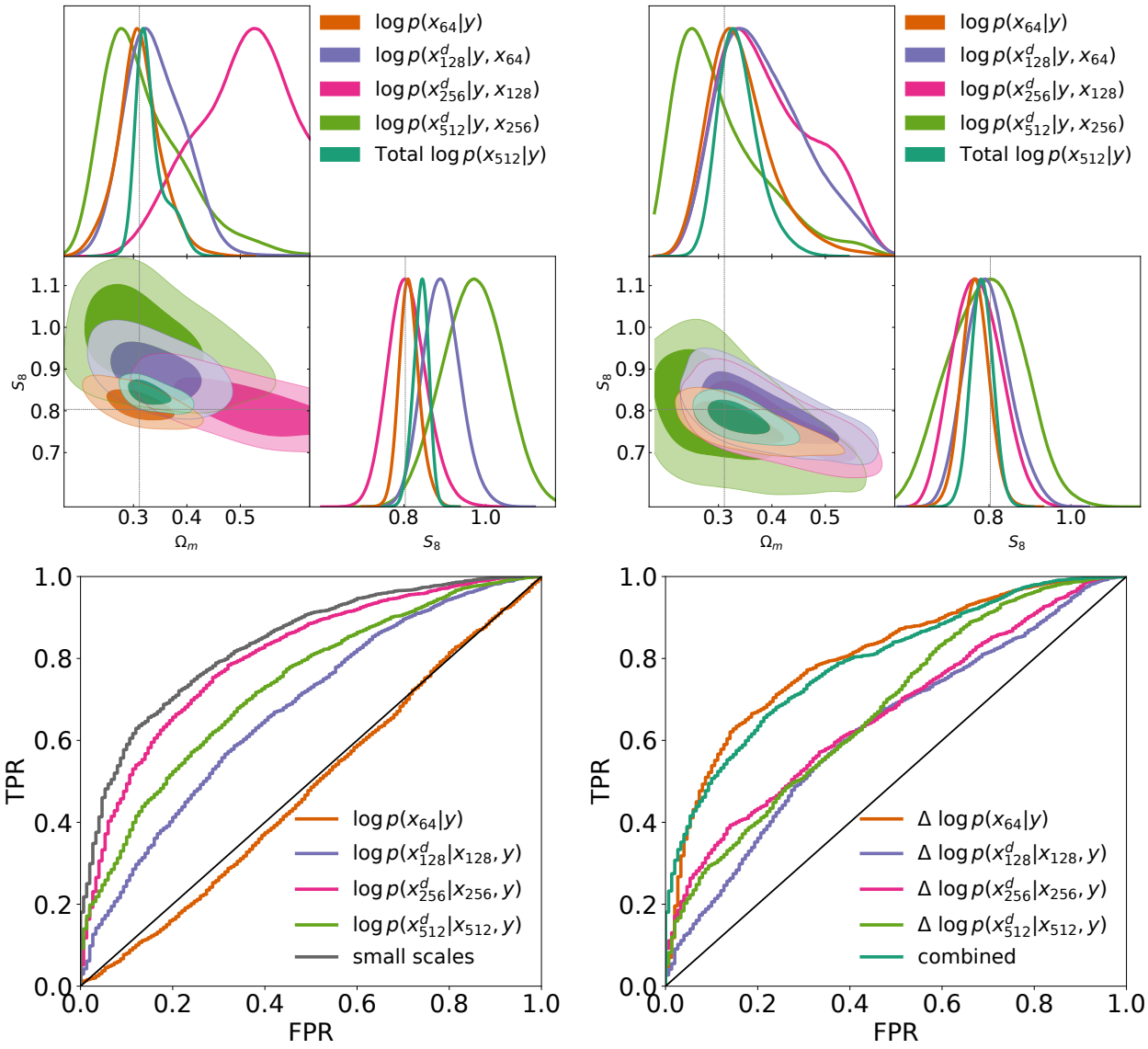



Figure 8.5: Top panel: scale-dependent posterior analysis of a baryon-corrected convergence map using Multiscale Flow trained on dark-matter-only maps (left), and Multiscale Flow trained on BCM maps (right). Bottom panel: ROC curve of identifying distribution shift with $\log p$ (left) and $\Delta \log p$ (right). The "small scales" in the lower left panel represent combining the three small scale terms. In these experiments, we consider $30\mathrm{arcmin}^{-2}$ galaxy shape noise.

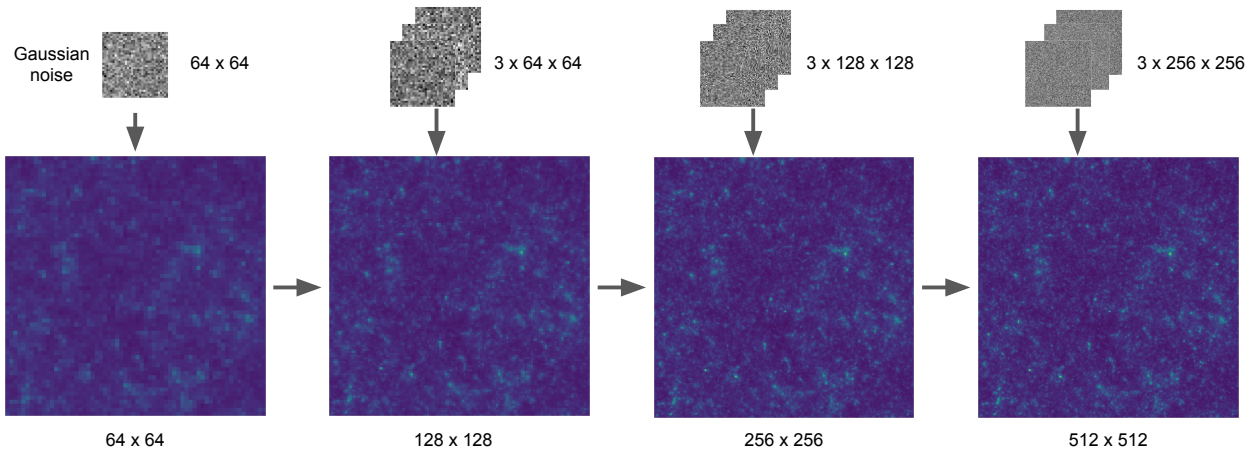



Figure 8.6: Illustration of Multiscale Flow sample generation (the reverse of Figure 8.1). The sample of the lowest resolution is first generated, and then small-scale information is gradually added. This process can also be viewed as super-resolution.

data. If we remove the small-scale information (because we believe the large scales are less likely to be affected by systematics), we can recover an unbiased constraint of cosmological parameters (orange contour). As a comparison, in the upper right panel, we show the posteriors from Multiscale Flow trained using maps with baryon physics. There is no distribution shift in this case and the information from the different scales is consistent.

In the bottom panel of Figure 8.5, we show the ROC curve of identifying this distribution shift with $\log p$ and $\Delta \log p$. As expected, the likelihood of large-scale term $\log p(x_{64}|y)$ cannot tell the difference between with and without baryon physics, while the likelihood of small-scale terms can be used for detecting the shifts. By combining all the small-scale terms, we get the best performance with AUROC of $0.84$. We also find that $\Delta \log p$ work equally well in this task. In this case the large-scale term $\Delta \log p(x_{64}|y)$ achieves the best performance with AUROC of $0.80$, because the small-scale constraints bias $y_{\rm MAP}$ away from $y_{\rm MAP,64}$. The two methods are essentially independent, and combining them further improves OoD detection. These maps have a small area ($3.5 \times 3.5 {\rm deg}^2$), and the 2048 test data used in this experiment span a wide range of baryon parameters, of which many are likely indistinguishable from the no baryons given the sampling variance between the maps. We expect our OoD detection methods will work even better for sky surveys with larger areas and for models where baryonic effects are more significant.

## Sample generation and super-resolution

We show an example of sample generation with Multiscale Flow in Figure 8.6. The process can also be viewed as iterative super-resolution of the low-resolution samples. In Figure 8.7 we show that Multiscale Flow samples and test data agree well in terms of the power spectrum and pixel

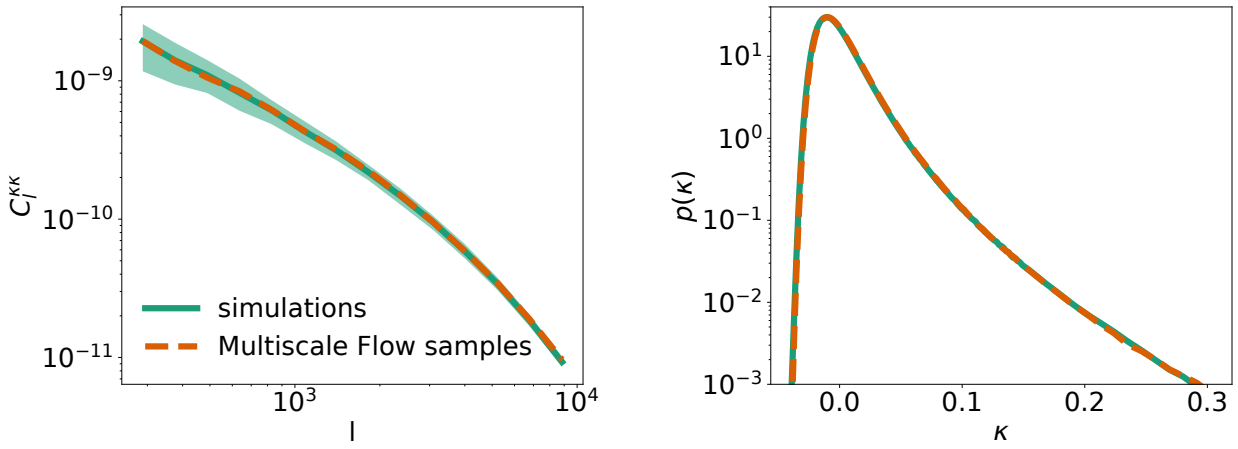



Figure 8.7: Comparison of the power spectrum (left) and pixel probability distribution function (right) between simulations and Multiscale Flow samples at fiducial cosmology.

probability distribution function. This demonstrates that Multiscale Flow samples can be used in lieu of expensive N-body simulations and ray tracing as a fast generator of mock data.

# Comparison with other machine learning models

## Comparison with discriminative models

So far there are lots of works using machine learning models to extract cosmological information at the field level. Most of these works either train models to directly learn the posterior constraints [333, 339], or build models to perform data compression $x \to s$ for cosmological inference, where the summary statistics $s$ can be a point estimate of the cosmological parameter [93, 92, 112, 270, 198, 197], or simply a data vector that contains rich information about cosmological parameter $y$ [202, 94]. These models are generally referred to as discriminative models.

Generative models, on the other hand, learn the data likelihood function $p(x|y)$, and then calculate the posterior distribution using Bayes rule. It has been suggested that while discriminative models have less asymptotic error, generative models have less sample complexity [235, 358, 366]. In other words, there can be two distinct regimes of performance as the training set size is increased. When the training size is small, the generative model achieves its asymptotic error much more quickly as data increases and can outperform the discriminative model, because the latter is more likely to overfit and requires more training data to converge.

For the weak lensing dataset considered in this work, the training set size is relatively small ($2.9 \times 10^4$ for maps without baryonic physics, and $7.7 \times 10^4$ for maps with baryonic physics) compared to the dimensionality $d = 512^2 \approx 2.6 \times 10^5$. This explains why Multiscale Flow, which learns the data likelihood function, outperforms CNN in Table 8.1 and 8.2. This explanation is further supported by the observation that Multiscale Flow never overfits when trained with

generative loss, and there is only slight overfitting when trained using hybrid loss with a large $\lambda$, which can be easily controlled with early stopping. The CNN training, on the other hand, overfits more easily due to its high sample complexity and requires more regularization techniques. In the future, we plan to investigate this topic more thoroughly and perform a detailed comparison of the two approaches with varying training sizes.

The low asymptotic error of discriminative models and low sample complexity of generative models can be understood as a bias-variance trade-off. To achieve the optimal balance of the trade-off, several works have proposed building hybrid models [263, 213, 32, 190]. Multiscale Flow is essentially a hybrid model, trained with a weighted combination of the generative loss (Equation 8.10) and discriminative loss (Equation 8.11). The interpolation parameter $\lambda$ balances the tradeoff of two approaches.

Apart from low sample complexity, another advantage of learning the likelihood function is robustness. The likelihood value itself contains information about whether the data may be contaminated by unknown systematic effects. As shown in the lower right panel of Figure 8.5, by comparing the likelihood value of a given data to those of the training data, we can tell whether the data is an outlier. It has also been suggested that generative models and hybrid models are more robust to adversarial attacks [185, 190], which could bias the parameter inference [136]. In the future, we plan to study more on making robust constraints against systematic effects.

## Comparison with diffusion models

Diffusion models have been shown to generate realistic astrophysical fields [294, 226, 365], and to achieve state-of-the-art performance on image density estimation tasks [160]. However, they seem to have difficulty producing reliable posterior constraints [52]. After all, the posterior is determined by the difference of log-likelihood across different conditional parameters, not the averaged log-likelihood. It has been suggested that different metrics (e.g., well-calibrated posterior v.s. realistic samples) are largely independent of each other in high dimensions, and good performance on one criterion does not imply good performance on other criteria [318]. In our experiments, we find that optimizing the model only with log-likelihood is not enough to produce reliable posteriors, due to the high asymptotic error of generative models. We train the model with hybrid loss to reduce the asymptotic error, which requires sampling the posterior during training with HMC. Considering that diffusion models are computationally too expensive to run HMC on the fly, we choose normalizing flows in this work.

## Comparison with TRENF

Translation and Rotation Equivariant Normalizing Flow has been shown to produce reliable and tight posterior constraints on Gaussian random field and mildly nonlinear matter density fields [56]. However, when we apply TRENF to weak lensing datasets in this work, it couldn't produce well-calibrated posteriors due to the restricted architecture. This motivates us to develop Multiscale Flow with affine coupling transforms [71, 161], which is able to approximate any probability distributions under mild conditions [172]. Moreover, the multiscale decomposition of the likelihood enables

scale-dependent posterior analysis that helps to detect domain shifts between training simulations and observed data.

## 8.5 Discussion

In this paper, we develop a Multiscale Flow model for field-level cosmological inference. Multiscale Flow tries to model the likelihood function of the cosmological field without any dimension reduction. If the field is learned perfectly, the resulting likelihood analysis becomes optimal. On mock weak lensing convergence dataset we demonstrate that the constraining power of Multiscale Flow outperforms the power spectrum in terms of Figure of Merit by factors of 2.5 - 4, depending on the noise level, and outperforms CNN by a factor 2 for the most realistic case with a noise level of $20\mathrm{arcmin}^{-2}$ and with baryon marginalization.

Multiscale Flow enables field-level scale-dependent posterior analysis, which helps the identification of scale-dependent systematics that are not accurately modeled in training simulations. We demonstrate that it is able to identify distribution shifts on weak lensing maps with baryonic physics if the model is trained with dark-matter-only maps.

In this paper, our main focus is optimal and robust field-level likelihood analysis, but we also show that Multiscale Flow can be used for fast sample generation and super-resolution, replacing the need for expensive N-body simulations and ray tracing. We expect many other applications of Multiscale Flow, such as 21cm and other intensity maps, weak lensing maps, projected galaxy clustering, X-ray and thermal SZ maps, etc. Multiscale Flow can also be used to model 3D galaxy fields or 1D spectrum data like Lyman alpha forest.

Multiscale Flow can be generalized to model maps with multiple channels $x_{2^n} = \{x^c_{2^n}\}_{c=1}^C$, where $c$ represents the index of channels, and $C$ is the total number of channels. Here the channels could represent different tomographic bins of cosmic shear analysis, or different tracers on the same area of the sky, such as galaxies and weak lensing. We can still use Equation 8.5 to decompose the likelihood of input maps with multiple channels, and each term can be further decomposed with

$$\begin{aligned}\log p(x_{2^k}|y) = &\quad \textstyle\sum_{c=1}^C \log p(x^c_{2^k}|x^1_{2^k},\cdots,x^{c-1}_{2^k},y),\\ \log p(x^d_{2^m}|x_{2^m},y) = &\sum_{c=1}^{C} \quad \log p(x^{d,c}_{2^m}|x^{d,1}_{2^m},\cdots,x^{d,c-1}_{2^m},x_{2^m},y),\end{aligned}$$

which allows us to check for consistency between different channels.

Multiscale Flow can also be generalized to model maps with survey masks. Following the strategy developed in [56], one can first sample noise at the masked region, and then introduce position-dependent flow transformation to the model to learn the effect of survey mask. It can thus be applied to realistic surveys such as Hyper Suprime-Cam [4], Euclid [179], or Vera C. Rubin Observatory Legacy Survey of Space and Time [141] for their robust and optimal analysis.

## 8.6 Appendix A: Materials and Methods

### Dark-matter-only weak lensing maps

The weak lensing convergence maps from Gupta et al.[112] are generated from a suite of 75 N-body simulations with spatially flat ΛCDM cosmologies. Each simulation differs in cosmological parameters $\Omega_m$ and $\sigma_8$, while the other cosmological parameters are fixed at $\Omega_b = 0.046$, $h = 0.72$ and $n_s = 0.96$. The two cosmological parameters $\Omega_m$ and $\sigma_8$ are sampled non-uniformly with density increases towards $\Omega_m = 0.26$ and $\sigma_8 = 0.8$. Each simulation evolves $512^3$ dark matter particles in a $240h^{-1}\mathrm{Mpc}$ box with N-body code gadget-2 [301]. A series of snapshots are saved between redshifts $0 < z < 1$ such that adjacent snapshots are separated by $80h^{-1}\mathrm{Mpc}$ in comoving distance.

Weak lensing convergence maps with field of view $3.5 \times 3.5\ \mathrm{deg}^2$ are then generated by ray-traced the snapshots of N-body simulations to redshift $z = 1$ with multiple lens plane algorithm [285]. 512 pseudo-independent maps are created from each simulation by randomly rotating, flipping, and shifting the simulation snapshots. We refer the reader to Gupta et al.[112] for a detailed description of how these data were generated.

Following Ribli et al.[270], we downsample the maps from resolution $1024^2$ to resolution $512^2$ ($\sim 0.4$ arcmin), and add Gaussian galaxy shape noise to the maps with a standard deviation

$$\sigma = \frac{\sigma_\epsilon}{\sqrt{2 n_{\mathrm{gal}} A_{\mathrm{pixel}}}}, \tag{8.16}$$

where $\sigma_\epsilon = 0.4$ is the mean intrinsic ellipticity of galaxies and $A_{\mathrm{pixel}}$ the area of the pixel. For this dataset we consider three different galaxy densities: $n_g = 10\ \mathrm{arcmin}^{-2}$, $n_g = 30\ \mathrm{arcmin}^{-2}$ and $n_g = 100\ \mathrm{arcmin}^{-2}$. Ribli et al.[270] smooth the maps with a 1 arcmin Gaussian kernel to increase the signal-to-noise (S/N) ratio and removes the information at very small scales where baryonic physics alters the matter distribution. In our analysis, however, we do not smooth the noisy maps. This is because our Normalizing Flow models the likelihood function by mapping the convergence map to a Gaussian random field of the same dimensionality, implicitly assuming that the input map is full-ranked. With Gaussian smoothing, the small-scale modes of smoothed maps become degenerate and the probability distribution is no longer full-ranked, leading to model failure in our analysis.

### Weak lensing maps with baryon

To study the impact of baryonic effects in our analysis, we also consider weak lensing convergence maps from Lu et al.[198]. These maps are generated from the same set of N-body simulations and ray-tracing algorithms as the dark-matter-only maps described above, and have the same resolution and field of view. The main difference is that the simulation snapshots are post-processed to include the baryonic effects. We briefly describe this post-processing step below and refer the reader to Lu et al.[196, 198] for more details.

Lu et al.[198] find all dark matter halos with mass $> 10^{12} M_\odot$ in the simulation snapshots, and replace the halo particles with spherically symmetric analytical halo profiles to characterize the

matter distribution inside halos. The analytical halo profile is given by Baryon Correction Model [BCM, 11], which describes the halos with four components: the central galaxy (stars), bounded gas, ejected gas (due to AGN feedback), and relaxed dark matter. The masses and profiles of these four components are parametrized by four free parameters: $M_c$ (the characteristic halo mass for retaining half of the total gas), $M_{1,0}$ (the characteristic halo mass for a galaxy mass fraction of $0.023$), $\eta$ (the maximum distance of the ejected gas from the parent halo), and $\beta$ (the logarithmic slope of the gas fraction vs. the halo mass). This post-processing step removes the substructure and non-spherical shape of the halos, but it has been shown that these morphological differences between the simulated halos and spherical analytical profiles are not statistically significant when compared to the uncertainties of the power spectrum and peak counts in an HSC-like survey [196].

Lu et al.[198] create 2048 maps with different baryon parameters for each cosmology. They train CNN with the first 1024 maps, and use the other 1024 maps to measure the mean and covariance matrix of the learned statistics. In our analysis, we only use the first 1024 maps to train our Multiscale Flow and do not use the rest of the 1024 maps.

Similar to the preprocessed steps of the dark-matter-only maps (described in the previous section), we downsample the maps to resolution $512^2$, and add Gaussian shape noise (Equation 8.16). For this dataset we consider four galaxy densities: $n_g = 10$ arcmin$^{-2}$, $n_g = 20$ arcmin$^{-2}$, $n_g = 50$ arcmin$^{-2}$ and $n_g = 100$ arcmin$^{-2}$, to better compare our results with Lu et al.[198].

## Multiscale Flow Hyperparameters

We use $p = 12$ block flows to model the large-scale term $\log p(x_{64}|y)$, and $q = 4$ block flows to model each of the three small-scale terms. The CNN in Equation 8.8 is chosen to be a convolutional residual neural network with 2 residual blocks and 64 hidden channels in the residual blocks.

## Summary Statistics Analysis

In this paper, we compare the performance of Multiscale Flow with analysis based on summary statistics. We consider not only standard summary statistics such as power spectrum and peak count, but also novel statistics such as scattering transform and convolutional neural networks (CNNs).

### Power Spectrum

We compute the power spectrum of the convergence maps using the publicly available LensTools package [250]. The power spectrum is calculated in 20 bins in the range $100 \leq l \leq 37500$ with logarithmic spacing, following the settings adopted in Ribli et al.[270] and Cheng et al.[44]. We take the logarithm of the power spectrum to be observable for parameter inference.

### Peak Count

Peak count has been widely used in current weak lensing analysis [209, 120, 370, 194]. In Table 8.1, we take the peak count measurement from Ribli et al. [270], who identify the local maxima

of convergence maps and measure the binned histogram of the peaks as a function of their $\kappa$ value. They use 20 linearly spaced $\kappa$ bins in total.

#### Scattering transform

Originally proposed by Mallat[203] as a tool to extract information from high-dimensional data, scattering transform has recently been applied to cosmological data analysis and shown improvement over the power spectrum in low noise regime [e.g., 44, 5, 43, 324]. For a given input field, the scattering transform first generates a group of new fields by recursively applying wavelet convolutions and modulus. The expected values of these fields are then defined as the scattering coefficients and used as the summary statistics. In this paper we compare our results directly to Cheng et al.[44], who estimate the constraining power of scattering transform using Fisher forecast on the same dataset.

#### Convolutional Neural Networks (CNN)

Several studies have explored using CNNs to construct summary statistics for cosmological inference [93, 202, 112, 149, 270, 198, 197]. In this work we compare our results on dark-matter-only weak lensing maps with Ribli et al.[270], and compare our results with Lu et al.[198] on weak lensing maps with baryons. Ribli et al.[270] and Lu et al.[198] train CNNs to predict cosmological parameters from the same convergence maps used in this work. Then they view these predicted parameters as summary statistics, and build Gaussian likelihood on these statistics for inference.

## Acknowledgements

We thank the Columbia Lensing group (`http://columbialensing.org`) for making their simulations available. B.D. thanks Xiangchong Li for helpful discussions on wavelet transform. This work is supported by U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research under Contract No. DE-AC02-05CH11231 at Lawrence Berkeley National Laboratory to enable research for Data-intensive Machine Learning and Analysis.

Table 8.2: Similar to Table 8.1, with baryonic effects.

| | Method | $10\text{arcmin}^{-2}$ | $20\text{arcmin}^{-2}$ | $50\text{arcmin}^{-2}$ | $100\text{arcmin}^{-2}$ |
|---|---|---|---|---|---|
| Fix baryon parameters at fiducial values | Multiscale Flow $p(x_{512}\|y)$ | **104** | **203** | **469** | **787** |
| | Multiscale Flow $p(x_{256}\|y)$ | 99 | 186 | 408 | 654 |
| | Multiscale Flow $p(x_{128}\|y)$ | 87 | 155 | 319 | 471 |
| | Multiscale Flow $p(x_{64}\|y)$ | 68 | 112 | 210 | 306 |
| | power spectrum | 41(41) | 61 (58) | 95(87) | 127(111) |
| | CNN | - | ($\sim$ 93) | ($\sim$ 146) | ($\sim$ 194) |
| Marginalize over baryon parameters | Multiscale Flow $p(x_{512}\|y)$ | **84** | **144** | **254** | **359** |
| | Multiscale Flow $p(x_{256}\|y)$ | 82 | 137 | 242 | 338 |
| | Multiscale Flow $p(x_{128}\|y)$ | 71 | 118 | 206 | 290 |
| | Multiscale Flow $p(x_{64}\|y)$ | 59 | 91 | 154 | 210 |
| | power spectrum | 34(33) | 48(48) | 68(65) | 84 (78) |
| | CNN | - | ($\sim$ 77) | ($\sim$ 109) | ($\sim$ 136) |

1. When fixing the baryon parameters at fiducial values, the FoM of CNN are estimated from Lu et al. [198]. Lu et al. [198] estimated the $1\sigma$ area of a $1500\text{deg}^2$ survey, and we scale their results by the area ratio for a direct comparison with our experiments.

2. For marginalizing over baryon parameters, simply rescaling the results of Lu et al. [198] by the area ratio underestimates the constraining power of CNN, due to the prior bounds of baryon parameters. Instead, we estimate its FoM by $\frac{\text{FoM}_{\text{PS,marginal}}}{\text{FoM}_{\text{PS,fiducial}}}\text{FoM}_{\text{CNN,fiducial}}$.

Table 8.3: Empirical coverage probability of posteriors from different methods, after marginalizing over baryon parameters. We report the percentage of test data that falls within the $68\%$ confidence region and the $95\%$ confidence region. A perfectly calibrated posterior should have $68\%$ and $95\%$ test data that fall in these two regions, respectively.

| Method | $n_g = 10\text{arcmin}^{-2}$ | $n_g = 20\text{arcmin}^{-2}$ | $n_g = 50\text{arcmin}^{-2}$ | $n_g = 100\text{arcmin}^{-2}$ |
|---|---|---|---|---|
| Multiscale Flow $p(x_{512}\|y)$ | $67.2\%$, $93.0\%$ | $72.7\%$, $95.3\%$ | $71.1\%$, $96.9\%$ | $71.1\%$, $98.4\%$ |
| Multiscale Flow $p(x_{256}\|y)$ | $65.6\%$, $94.5\%$ | $71.1\%$, $95.3\%$ | $74.2\%$, $96.1\%$ | $73.4\%$, $97.7\%$ |
| Multiscale Flow $p(x_{128}\|y)$ | $68.0\%$, $94.5\%$ | $71.9\%$, $96.1\%$ | $75.0\%$, $95.3\%$ | $77.3\%$, $97.7\%$ |
| Multiscale Flow $p(x_{64}\|y)$ | $66.4\%$, $96.9\%$ | $67.2\%$, $96.1\%$ | $72.7\%$, $96.1\%$ | $73.4\%$, $97.6\%$ |

# Chapter 9

# A comparative study of cosmological constraints from weak lensing using Convolutional Neural Networks

The contents of this chapter were originally published in [291],

A comparative study of cosmological constraints from weak lensing using Convolutional Neural Networks

*Sharma D., Dai B., Seljak U.*(arXiv:2403.03490) (submitted to JCAP)

Weak Lensing (WL) surveys are reaching unprecedented depths, enabling the investigation of very small angular scales. At these scales, nonlinear gravitational effects lead to higher-order correlations making the matter distribution highly non-Gaussian. Extracting this information using traditional statistics has proven difficult, and Machine Learning based summary statistics have emerged as a powerful alternative. We explore the capabilities of a discriminative, Convolutional Neural Networks (CNN) based approach, focusing on parameter constraints in the ($\Omega_m$, $\sigma_8$) cosmological parameter space. Leveraging novel training loss functions and network representations on WL mock datasets without baryons, we show that our models achieve $\sim 5$ times higher figure of merit in the $\sigma_8$ - $\Omega_m$ plane than the power spectrum, $\sim 3$ times higher than peak counts, and $\sim 2$ times higher than previous CNN-learned summary statistics and scattering transforms, for noise levels relevant to Rubin or Euclid. For WL convergence maps with baryonic physics, our models achieve $\sim 2.3$ times stronger constraining power than the power spectrum at these noise levels, also outperforming previous summary statistics. To further explore the possibilities of CNNs for this task, we also discuss transfer learning where we adapt pre-trained models, trained on different tasks or datasets, for cosmological inference, finding that these do not improve the performance.

# 9.1 Introduction

Weak gravitational lensing (WL) is the distortion of light from distant galaxies caused by the gravitational influence of intervening large-scale structures that trace total matter in the universe. This phenomenon creates a subtle cosmic shear pattern in the sky, altering the observed shapes and orientations of background galaxies. The distortion of galaxy shapes, quantified through summary statistics, holds valuable information about underlying cosmological parameters [17, 159]. Various surveys, such as the Dark Energy Survey (DES), Hyper Suprime-Cam Survey (HSC), Euclid, the Vera Rubin Observatory (Rubin), and the Nancy Grace Roman Space Telescope (Roman), aim to map this cosmic shear across large areas of the sky. These surveys will provide observational data that can be used to constrain fundamental cosmological parameters, particularly $\Omega_m$ (matter density) and $\sigma_8$ (amplitude of matter fluctuations) that WL signals are most sensitive to in the standard cosmological model [153, 174, 131, 116].

To extract information from WL data, various summary statistics are employed. At the two-point level, WL data are analyzed using correlation functions such as the power spectrum of the shear or convergence. However, these traditional summary statistics leave a wealth of information untapped in the WL signal due to the highly non-Gaussian features at small scales. To address this limitation, various other summary statistics have emerged, such as higher-order correlation functions [308, 309, 284, 360, 245, 246, 286, 290, 98, 329, 330]. These are challenging because of the large number of coefficients, and the difficulty to measure the associated covariance matrix, as well as sensitivity to outliers [158]. Other methods that have been proposed are peak counts [206, 69, 176, 357, 155, 369], Minkowski functionals [220, 276, 111, 175, 331, 241], scattering transform coefficients [5, 44, 323, 100], and statistics learned by neural networks [93, 41, 202, 112].

Convolutional Neural Networks (CNNs) have been employed in various studies to estimate cosmological parameters from WL convergence maps [112, 270, 93, 92, 334, 149, 202, 198, 197]. [112] trained CNNs on noise-free convergence maps, showcasing a 5 times improvement in the precision of $\Omega_m - \sigma_8$ constraints compared to the power spectrum analysis. [270] trained CNNs with a different architecture on convergence maps with different levels of shape noise, achieving a 2.4–2.8 times improvement over power spectrum in parameter constraints for surveys like Rubin. [92] leveraged the KiDS-450 tomographic WL dataset, illustrating a 30% enhancement in $S_8 = \sigma_8(\Omega_m/0.3)^{0.5}$ parameter constraints compared to power spectrum analysis. [210] reported a 20% improvement in constraints over traditional methods (power spectrum, peak counts, and Minkowski functionals) using their CNN framework. [149] [198] studied the impact of baryonic effects on WL analysis with CNNs, and they further applied their framework to HSC first-year data, finding a factor of 3 improvement in $\Omega_m$ constraints over power spectrum [197]. [148] trained CNNs on Dark Energy Survey (DES) Year 3 weak-lensing maps and illustrated more than a factor of 2 improvement in $S_8$ - $\Omega_m$ constraints compared to power spectrum inference.

More recently, [56] and [55] proposed using generative Normalizing Flows to create samples and model the field-level likelihood of weak lensing maps. On weak lensing mock datasets, [55] showed that Multiscale Flow (MSF) outperforms the power spectrum by factors of $3 - 9$ for different noise levels relevant to different surveys. It also achieves about two times higher constraining power

when compared to peak counts, CNNs, and scattering transform summary statistics. However, CNNs are easier to train and work with than the generative Multiscale Flow model. Moreover, for cosmological inference from surveys with large survey areas, CNNs can be applied effectively by cropping the survey images and combining the summary statistics obtained from the CNNs [197]. Multiscale Flow, being a generative model with full dimensionality of the data, does not scale as well with growing survey size.

The ability of MSF to outperform current implementations of CNN raises the question whether this is inherently due to their generative training, as has been suggested in previous work [235]. An alternative explanation is that the existing CNN analyses have not been optimal in terms of architectural choices and training methods. In this paper, we therefore explore the ability of CNNs to constrain cosmology using a variety of novel techniques and loss functions that have been recently proposed. We train our models on WL convergence maps with and without baryons and compare their performance to that of the power spectrum, peak counts, scattering transform and previous CNN works. We will show that our models outperform these statistics by considerable factors.

This paper is organized as follows: Section 9.2 provides a comprehensive overview of the maps we used to train and test our models, the methodology employed during training, and post-training-prediction procedures to obtain parameter constraints. In Section 9.3, we present and discuss our results of parameter constraints derived from the trained models. Section 9.4 explores our results employing transfer learning, where we adapt pre-trained models to predict cosmologies using the same WL maps. We summarize and conclude in section 9.5.

## 9.2 Materials and Methods

In this section, we describe the data, models, and methods we use to get stringent parameter constraints in $\Omega_m - \sigma_8$ cosmological plane. First, we describe the weak lensing convergence maps that we use for training and testing our models. We use two different datasets for the types of convergence maps: one with only dark matter, and another dataset with baryonic physics added. Second, we describe the CNN models used and the training loss functions and methodologies we implemented during training to get varying degrees of constraining powers. Finally, we describe how we use our trained CNNs' outputs as a summary statistic for parameter inference as well as the other summary statistics that we compare our results against.

### Weak lensing maps

#### Dark-matter-only weak lensing maps

The dark-matter-only (DM-only) weak lensing convergence maps utilized in this study, obtained from [112], stem from a suite of 80 N-body simulations characterized by spatially flat ΛCDM cosmologies. Each simulation varies in cosmological parameters $\Omega_m$ and $\sigma_8$, while maintaining fixed values for other parameters: $\Omega_b = 0.046$, $h = 0.72$, and $n_s = 0.96$. Specifically, the sampling of $\Omega_m$ and $\sigma_8$ is non-uniform, concentrating more densely around $\Omega_m = 0.26$ and $\sigma_8 = 0.8$. These simulations evolve $512^3$ dark matter particles within a $240h^{-1}$ Mpc box, employing the N-body

code Gadget-2 [301]. Snapshots are recorded between redshifts $0 < z < 1$, spaced $80h^{-1}$ Mpc apart in comoving distance.

Weak lensing convergence maps with a field of view of $3.5 \times 3.5$ deg$^2$ are then generated by ray-tracing the snapshots of N-body simulations to redshift $z = 1$ with a multiple lens plane algorithm [285]. From each simulation, $512$ pseudo-independent maps are derived by incorporating random rotations, flips, and shifts to the snapshots. Following the setup from [112], all the source galaxies are placed at redshift z = 1, and tomography is not considered. Detailed insights into the map generation process can be found in [112]. In this work our primary goal is to understand how much information can be extracted at the field level with CNNs, and how it compares with various summary statistics. Therefore we directly take the simulated maps with simple setups from previous works for a direct comparison. We plan to study more realistic setups and systematic effects such as tomographic analysis and intrinsic alignment in our future works.

The convergence maps from [270] have a resolution of $1024^2$. We first downscale the maps, using average pooling of every $4\times4$ patch, to $256^2$ for faster training times, without seeing considerable degradation in the constraining power of the trained models. Then, we add Gaussian shape noise to the $256^2$ resolution maps. The standard deviation $\sigma$ of the shape noise is computed as

$$\sigma = \frac{\sigma_\epsilon}{\sqrt{2 n_{\rm g} A_{\rm pixel}}} \tag{9.1}$$

with $\sigma_\epsilon \sim 0.4$ denoting the mean intrinsic ellipticity of galaxies, and $A_{\rm pixel}$ signifying the pixel area. This dataset incorporates three distinct galaxy density scenarios: $n_{\rm g} = 10$, 30, and 100 arcmin$^{-2}$. A noise level of 10 galaxies arcmin$^{-2}$ characterizes typical ground-based surveys like CFHTLens, DES, or KiDS. Around 30 galaxies arcmin$^{-2}$ represent the targeted noise level for surveys like Rubin or Euclid, while future space missions like Roman might access between 50 and 75 galaxies arcmin$^{-2}$. The scenario with 100 galaxies arcmin$^{-2}$ stands as an optimistic anticipation for forthcoming space-based surveys. While [270] apply a 1 arcmin Gaussian kernel for map smoothing to augment signal-to-noise ratio (S/N) and mitigate small-scale information affected by baryonic physics, we find no need for this procedure, and our analysis does not use smoothing techniques on the noisy maps.

**Weak lensing maps with baryons**

To understand the influence of baryonic effects within our analysis, we analyze weak lensing convergence maps from [198]. These maps stem from an identical set of N-body simulations and ray-tracing methodologies as the previous dark-matter-only maps, with identical resolution and field of view. However, a key distinction lies in the post-processing of simulation snapshots to incorporate baryonic effects. A concise overview of this post-processing step follows herein; for more comprehensive details, readers are referred to [198, 196].

In their work, [198] identify all dark matter halos with masses greater than a mass threshold of $10^{12} M_\odot$ within simulation snapshots. These halos' constituent particles are then substituted with spherically symmetric analytical halo profiles, characterizing the matter distribution within the halos. The analytical halo profile, derived from the Baryon Correction Model (BCM) [11],

delineates halos through four components: the central galaxy (stars), bounded gas, ejected gas (attributed to AGN feedback), and relaxed dark matter. Parameterized by four free parameters – $M_c$ (characteristic halo mass retaining half the total gas), $M_{1,0}$ (characteristic halo mass for a galaxy mass fraction of $0.023$), $\eta$ (maximum distance of the ejected gas from the parent halo), and $\beta$ (logarithmic slope of the gas fraction vs. the halo mass). This model eliminates the substructure and non-spherical shapes of the halos, but it has been argued that the morphological disparities between simulated halos and spherical analytical profiles are statistically insignificant compared to uncertainties in the power spectrum and peak counts within an HSC-like survey [198].

[198] generate 2048 maps for each cosmology, each with distinct baryon parameters. They utilize the first 1024 maps for training a CNN and employ the remaining 1024 maps to calculate the mean and covariance matrix of the CNN-learned summary statistics. Aligned with their methodology, our analysis also uses the first 1024 maps for training and the subsequent 1024 maps for statistical inference. Similar to the preprocessing steps applied to the dark-matter-only maps, we first downsample these maps to a resolution of $256^2$, using average pooling, and then introduce Gaussian shape noise. This dataset encompasses three galaxy densities: $n_g = 20$, $50$, and $100$ arcmin$^{-2}$, facilitating a comprehensive comparison with the results of [198].

## Neural network architectures

In our study, we leverage the learning power of deep residual neural networks (ResNets) [122, 123]. Specifically, we experimented with ResNet18, ResNet34, and ResNet50. These architectures have played a pivotal role in image data based deep learning over the past several years. ResNet18, ResNet34, and ResNet50 are characterized by their depth, a defining feature that allows them to capture intricate patterns and nuances in the data. The innovative element of ResNets is the introduction of residual connections, or skip connections, which facilitate the training of very deep networks. This helps overcome the vanishing gradient problem and, as a result, enables the training of neural networks with many layers. ResNet18, the smallest of the three, consists of 18 convolution layers, ResNet34 is deeper with 34 layers, with ResNet50 being even deeper, featuring 50 layers. These networks have demonstrated remarkable performance across a wide range of computer vision tasks, making them invaluable tools for image based data such as weak lensing data.

Our empirical experiments show that the choice of neural network architecture does not significantly influence our results. We present a comparison between different neural network architectures using the same loss function in table 9.1. Hence, we opted to employ ResNet18 consistently for all experiments detailed in this paper. We made this choice due to its relatively shallower depth compared to ResNet34 and ResNet50, balancing computational complexity while preserving the capacity to capture meaningful representations from our weak lensing data.

## Training

### Training Loss Functions

We utilized various loss functions to train our models, aiming to optimize the performance of our networks in capturing and interpreting the complex cosmological parameter space. We detail the loss functions in the following paragraphs. We find that CNNs achieve different constraining power depending on the choice of the loss function, as detailed in section 9.3.

First, we employed the Mean Squared Error (MSE) loss, defined as:

$$\text{MSE loss} = \frac{1}{n}\sum_{\text{i=1}}^{n}(y_i - \hat{y}_i)^2 \tag{9.2}$$

This loss function measures the squared differences between predicted ($\hat{y}_i$) and actual ($y_i$) values across $n$ data points. It corresponds to maximum likelihood estimation (MLE), under the assumption that the likelihood is Gaussian with an identity covariance matrix.

Another loss function we use involves transforming cosmological parameters into approximately uncorrelated parameters. Specifically, we introduced $S_8$ as a function of $\sigma_8$ and $\Omega_m$ via $S_8 = \sigma_8 \left(\frac{\Omega_m}{0.3}\right)^{0.5}$, coupled with an approximately orthogonal parameter Ortho $= \frac{\sigma_8^2}{2} - \Omega_m^2$. Using an L2 loss function with custom weights, corresponding to the expected errors associated with these parameters, we formulated the following loss:

$$\text{MSE}_{\text{NP}}\text{ loss} = w_1 \cdot (S_8 - S_{8,\text{true}})^2 + w_2 \cdot (\text{Ortho} - \text{Ortho}_{\text{true}})^2 \tag{9.3}$$

Here, $w_1$ and $w_2$ denote the expected inverse variance for $S_8$ and Ortho, respectively. This loss has the advantage of incorporating non-Gaussian correlations, due to the nonlinear nature of the parameter transformation.

Next, we explored Principal Component Analysis (PCA) by transforming the original parameter space into a different space with the same dimensionality (2 for DM-only maps and 6 for baryon maps), with the goal of optimizing training parameters for the MSE loss. This transformed space, characterized by principal components capturing data variations, allows us to use an L2 loss function:

$$\text{MSE}_{\text{PCA}}\text{ loss} = (\mathbf{X} - \hat{\mathbf{X}})^T \cdot \mathbf{W} \cdot (\mathbf{X} - \hat{\mathbf{X}}) \tag{9.4}$$

Here, $\mathbf{X}$ represents the parameter set in the transformed PCA space, $\hat{\mathbf{X}}$ indicates predicted values in the same space, and $\mathbf{W}$ symbolizes a diagonal weight matrix holding expected inverse variance values for each parameter. By utilizing PCA followed by an L2 loss, we improve model efficiency over the regular MSE loss. However, the PCA transformation is linear and unable to account for non-Gaussian posteriors.

Next, we integrated the Neural Posterior Estimation (NPE) methodology proposed by [149, 361] for Simulation-Based Inference (SBI). This method aims to directly learn the posterior distribution by employing a Neural Density Estimator (NDE) like a Normalizing Flow (NF). The loss function

in this context, called the Variational Mutual Information Maximization (VMIM) loss [14] or the expected Negative Log-Likelihood (NLL) loss ($L_{\rm NLL}$), minimizes the expected negative log-posterior:

$$\text{VMIM loss} = \mathbb{E}_{\mathrm{p}(\theta,\mathrm{x})}\left[-\log p_\phi(\theta|x)\right] = L_{\rm NLL} \tag{9.5}$$

This loss function involves compressing the WL maps using a neural network – a ResNet in our case – into a summary statistic, and employing conditional normalizing flows to approximate the posterior distribution $p_\phi(\theta|x)$ given learned summary statistic. In our work, we used a conditional normalizing flow based on RealNVP [71] with 4 layers to approximate the posterior distribution.

## Hyperparameters and Training Strategies

Throughout the training phase, we conducted careful experimentation with various hyperparameters to yield optimal model performance and convergence. To prevent overfitting and regularizing the model, we used a weight decay factor of $1 \times 10^{-3}$ using the AdamW optimizer with default settings for optional parameters such as $\beta_1 = 0.9$ and $\beta_2 = 0.999$. We established this value of the weight decay factor to be optimum after experimenting with a wide range of values. This factor played a crucial role in constraining the complexity of the model and preventing excessive sensitivity to noise in the training data.

As described in section 9.2, there are 512 DM-only WL maps per cosmology. Out of these, we used 360 maps as training data, and 152 maps were used for parameter inference post-training. Of the 152 maps used for parameter inference, 100 maps were used for calculating the mean and covariance matrix of the CNN-learned summary statistic, and 52 maps were used for testing the models and measuring the FoM. We removed the fiducial cosmological model from the training data and only used it for validation, in order to prevent overfitting to the training cosmology and test the accuracy of model interpolation between different cosmological models.

For WL maps with baryons, as described in section 9.2, we used 1024 maps per cosmology as training data and the remaining 1024 maps per cosmology to calculate the mean and covariance matrix of the CNN-learned summary statistics. For measuring the FoM, we used an additional 128 maps at fiducial parameters. As in the case of DM-only maps, we removed the maps at fiducial cosmology from training data and only used them for testing.

Additionally, we organized the training data into batches during training, with each batch containing randomly sampled 128 WL maps. This batch size was selected to balance computational efficiency and model convergence while ensuring a manageable memory during training.

In our training process, a piecewise constant learning rate schedule was adopted to dynamically adjust the learning rate at specific milestones throughout the training iterations. This adaptive learning rate schedule was essential in controlling the rate of model parameter updates, enhancing the model's ability to navigate the loss landscape effectively and converge toward an optimal solution.

Despite the use of regularization, we found ResNets to overfit the training data. To further mitigate overfitting, we employed early stopping. As stated above, we separated fiducial cosmology

maps from the training data and designated them for validation purposes. This ensures that the model does not overfit since the model is not directly trained on fiducial cosmology maps. In order to fix any bias in the CNN training, we treat our model outputs as summary statistics, using maps with different realizations to calculate its mean and covariance, for parameter inference as outlined in section 9.2. During training iterations, the model's performance on the validation set was monitored, and the model achieving the lowest loss on this validation set was saved. This approach ensured that the model's generalizability was preserved by preventing it from becoming overly specialized to the training data and thereby improving its performance on unseen test data.

## Parameter Inference

Neural network output predictions for cosmological parameters can be interpreted as summary statistics, akin to conventional statistics like the power spectrum or peak counts, regardless of their direct connection to the underlying parameters. This approach has been widely applied in prior studies [112, 270, 210, 198, 197]. Hence, a statistic, in this context, denotes either the prediction output by a network or a combination of various statistics. The advantage of this summary statistic view is that it can be further improved if necessary, and that its density estimation is relatively simple since it is very low dimensional. In contrast, high dimensional summary statistics suffer from the need to have many simulations to learn their covariance matrix, and typically one needs several times more training data than the number of summary statistics.

Following previous studies, given a summary statistic comprising $d$ observables, we model the likelihood of observing $y$ as a multidimensional Gaussian distribution:

$$p(y|\theta) \propto \frac{1}{\sqrt{\det C}} \exp\left(-\frac{1}{2}[y - y(\theta)]^T C^{-1}(\theta)[y - y(\theta)]\right). \tag{9.6}$$

Here, $\theta$ represents the underlying cosmological and baryonic parameters, and we use $N$ realizations to estimate the covariances $C$. For calculating the expected values, $y(\theta)$, and the covariances, $C(\theta)$, we follow slightly different approaches for the two different WL map datasets, which the following subsections outline.

### Dark Matter Only Maps

To extend the likelihood across the entire $\Omega_m - \sigma_8$ space from the 80 discrete cosmologies sampled by the DM-only maps, we perform linear interpolation for both the expected value, $y(\theta)$, and the covariances. Using Bayes' theorem, we then estimate the posterior distribution of $\theta$ considering a mock observation at the fiducial parameters $y(\theta_0)$ while employing uniform priors on all parameters.

### Baryon Maps

To extend the likelihood across the $\Omega_m - \sigma_8$ space from the 75 discrete cosmologies sampled by the baryon maps, we undertake a two-step interpolation for the expected value $y(\theta)$ following [198]: (1) fitting a second-degree polynomial for the four baryonic parameters within each cosmology,

and (2) linear interpolation between cosmologies. We assume constant covariances across different baryonic parameters, estimating these covariances for each cosmology and interpolating them linearly. Using Bayes' theorem, we estimate the posterior distribution of $\theta$ based on a mock observation at the fiducial parameters $y(\theta_0)$, while adopting log-uniform priors on all parameters.

### Power Spectrum

We take the power spectrum constraining power estimation from [55], who measure the power spectrum of convergence maps using the publicly available LensTools package [250]. The power spectrum is calculated in 20 bins in the range $100 \leq l \leq 37500$ with logarithmic spacing, following the settings adopted in [270] and [44]. The logarithm of the power spectrum is taken to be observable for parameter inference.

### Peak Count

For peak count constraining power, we take the peak count measurement from [270], who identify the local maxima of convergence maps and measure the binned histogram of the peaks as a function of their $\kappa$ value. They use 20 linearly spaced $\kappa$ bins in total.

### Scattering Transform

Initially introduced as a method to extract information from high-dimensional data by [203], the scattering transform has been recently applied to cosmological data analysis. In studies [44, 5, 43, 324], it showcased improvements over the power spectrum, especially in low noise scenarios. This technique generates a cluster of new fields through recursive wavelet convolutions and modulus operations. The expected values of these derived fields serve as scattering coefficients and act as summary statistics. In this study, we compare our findings directly to the results obtained by [44], who estimated the constraining capabilities of the scattering transform using Fisher forecast techniques on the same dataset that we use.

## 9.3 Results

In this section, we present the constraining power of our trained CNN models for each loss mentioned above and compare them to traditional summary statistics as well as previous CNN-learned summary statistics. First, we describe the results for the DM-only WL maps dataset, and then we explore the effects of baryons on our constraining power using the baryon maps dataset.

We use the same simulated maps and follow most data processing and inference setups as the previous studies with which we compare our results. The only differences in our methods are that we downscale the maps to $256^2$ resolution and that we do not smooth the maps after adding shape noise. These differences do not significantly change our results, as shown in table 9.2. We measured the FoM and 1D marginalized constraints using posterior samples. On the other hand,

[44] estimate the FoM of the scattering transform using the Fisher matrix, which is an upper limit of the true FoM according to the Cramér-Rao inequality.

## Cosmological constraints from noisy weak lensing maps

First, we evaluate the efficacy of our trained CNNs using the $3.5 \times 3.5$ deg$^2$ mock weak lensing convergence maps from [270] to conduct field-level inference. The performance of our model is assessed based on the constraining power, measured as the reciprocal of the $1\sigma$ confidence area on the ($\Omega_m$, $\sigma_8$) plane. Table 9.1 presents a comparative analysis showcasing the figure of merit of our CNNs across different noise levels, combined with results obtained using traditional summary statistics such as the power spectrum, peak count, scattering transform [44], and statistics derived from CNNs in [270].

Our models achieve the best performance among all the methods, surpassing the other methods, including the power spectrum, by substantial margins. Specifically, on galaxy densities $n_g = 10$, $30$ (relevant for Rubin or Euclid), and $100$ arcmin$^{-2}$, our CNNs outperform the power spectrum by factors of $2.4$, $5$, and $5$, respectively, using the VMIM loss. Notably, our best-performing model demonstrates a marked enhancement in constraining power, achieving $\sim 1.5 - 2$ times higher performance compared to peak counts, previous CNN models, and scattering transform for all noise levels.

Using $MSE_{\rm NP}$ and $MSE_{\rm PCA}$ losses, our CNNs also outperform power spectrum analysis by factors of $\sim 2$, $3$, and $4$ for the three noise levels respectively. They also achieve $\sim 1.2 - 2$ times stronger parameter constraints than peak counts, previous CNN models, and scattering transform for all noise levels. Using MSE loss, we achieve comparable constraining power to [270] who also used MSE loss, but used a different CNN architecture and applied smoothing to the fields. This suggests that the constraining power achieved by different deep models is fairly independent of model architectures, and it is the choice of the loss function that is more important for optimal performance.

To illustrate the impact of different losses on our model's performance, we present the posterior comparison of various losses using 16 test maps with galaxy number density of $n_g = 10, 30, 100$ arcmin$^{-2}$ in Figure 9.1. The posterior constraints are consistent with the true cosmological parameters (shown in black dashed lines) for all noise levels. In alignment with Table 9.1, the VMIM loss achieves the most stringent parameter constraints, followed by $\rm MSE_{NP}$, $\rm MSE_{PCA}$, $\rm MSE$ losses, in that order. Additionally, we present the 1D marginal constraints of $\Omega_m$, $\sigma_8$ and $S_8$ in table 9.3.

## Impact of baryons

To study the impact of baryons on parameter constraints achieved using our models, we apply CNNs on mock weak lensing maps incorporating baryonic physics. Table 9.4 presents the observed constraining power variations across diverse galaxy-shape noise levels. In the presence of baryons, our models have $\sim 2.3$ times higher constraining power on cosmological parameters than the power spectrum. Using the VMIM loss, our results also outperforms previous CNN results by a factor of

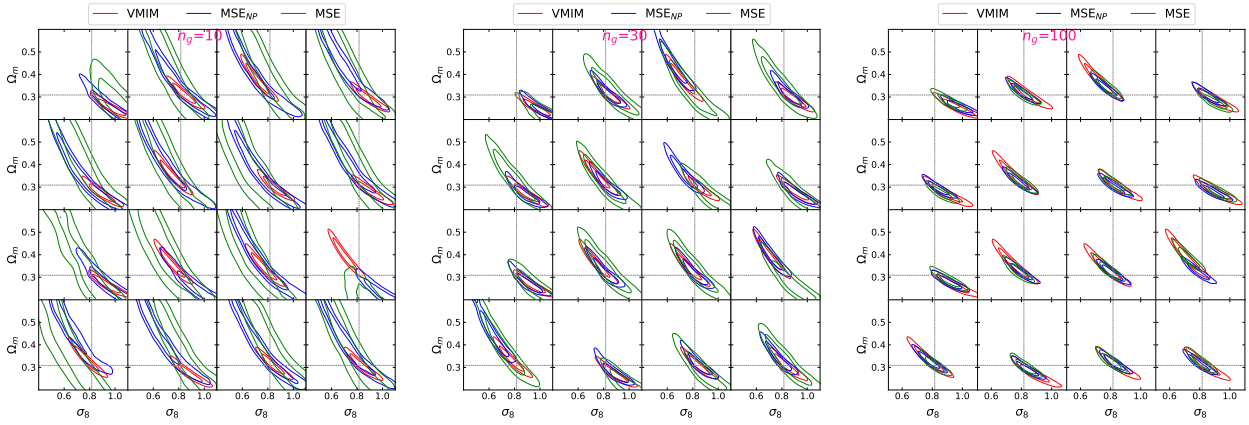



Figure 9.1: Comparison of the constraining power of different losses on 16 test data with the panels corresponding to galaxy number density $n_g = 10, 30, 100\,\text{arcmin}^{-2}$ respectively. The posterior constraints contain the true cosmological parameters (shown in black dashed lines) for all noise levels. In correspondence with Table 9.1, the VMIM loss exhibits the most stringent parameter constraints, succeeded by $MSE_{\text{NP}}$, $MSE_{\text{PCA}}$, and $MSE$ losses, respectively. The constraining power achieved by the MSE loss is comparable to the results of [270]. The VMIM loss achieves $\sim 2$ times smaller contours compared to the MSE loss and [270].

Table 9.1: Comparison of the constraining power between different methods at fiducial cosmology. The figure of merit is measured by the reciprocal of the $1\sigma$ confidence area on the $(\Omega_m, \sigma_8)$ plane, using a $3.5 \times 3.5\,\text{deg}^2$ convergence map. The rows with bold figures of merit correspond to results obtained using methods outlined in this study.

| Method | $n_g = 10\,\text{arcmin}^{-2}$ | $n_g = 30\,\text{arcmin}^{-2}$ | $n_g = 100\,\text{arcmin}^{-2}$ |
|---|---|---|---|
| ResNet18 + VMIM loss | 72 | 236 | 400 |
| ResNet34 + VMIM loss | 72 | 234 | 403 |
| ResNet50 + VMIM loss | 75 | 239 | 407 |
| ResNet18 + $\text{MSE}_{\text{NP}}$ loss | 70 | 170 | 351 |
| ResNet18 + $\text{MSE}_{\text{PCA}}$ loss | 50 | 140 | 290 |
| ResNet18 + MSE loss | 35 | 114 | 297 |
| Power Spectrum | 30 | 52 | 81 |
| Peak Count | 40 | 85 | 137 |
| CNN [270] | 44 | 121 | 292 |
| Scattering Transform | $\lesssim 50$ | $\lesssim 140$ | $\lesssim 329$ |
| Multiscale Flow ($256^2$) | 82 | 226 | 631 |

Table 9.2: Comparison of the constraining power using different map processing methodology at $n_g = 30$ arcmin$^{-2}$. We compare the figure of merit achieved by our models when they are trained on (1) maps of $256^2$ resolution without smoothing after adding shape noise (our default setup), (2) maps of $256^2$ resolution with smoothing, and (3) maps of $512^2$ resolution without smoothing. The difference between these setups is less than $10\%$ and is much smaller than the difference between different methods in Table 9.1.

| Method | No smoothing, $256^2$ | Smoothing, $256^2$ | No smoothing, $512^2$ |
|---|---|---|---|
| ResNet18 + VMIM loss | 236 | 217 | 249 |

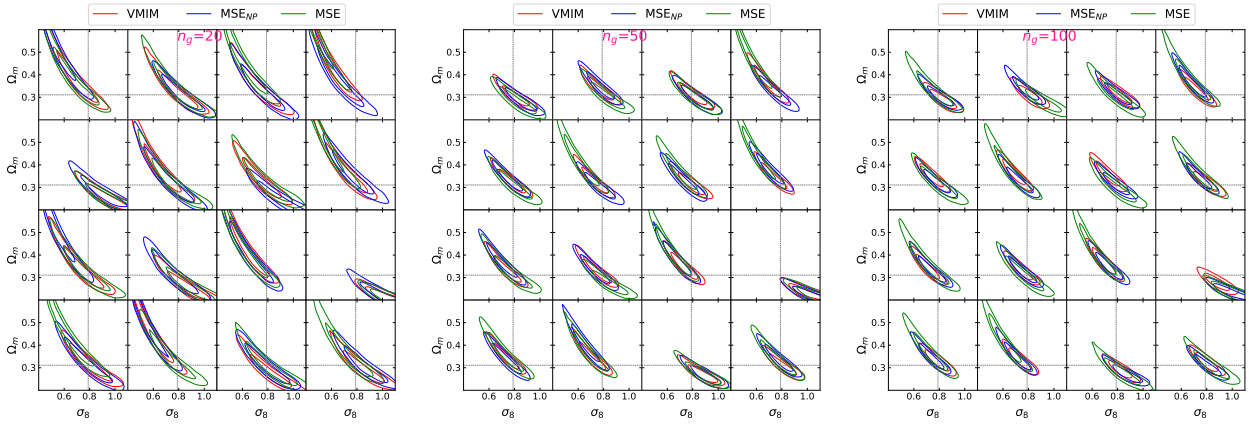



Figure 9.2: Comparison of the constraining power of different losses on 16 test data with the panels corresponding to galaxy number density $n_g = 20, 50, 100$ arcmin$^{-2}$ respectively in the presence of baryons.

$\sim 1.5$. We present the 1D marginal constraints of $\Omega_m$, $\sigma_8$, and $S_8$ in the presence of baryons in table 9.3.

Additionally, in Figure 9.2, we present the posterior comparison derived from assessing various loss functions using 16 test maps characterized by a galaxy number density of $n_g = 20, 50, 100$ arcmin$^{-2}$. This comparison highlights the impact of different loss functions on the model's performance and inference outcomes.

## 9.4 Transfer Learning

Transfer learning is a Machine Learning (ML) technique where insights acquired from one task are leveraged to enhance performance in a related task [34, 33, 367]. For example, a network that has been trained to identify cars in image classification could be utilized, after fine-tuning, when attempting to identify trucks. Due to its wide prospects of application in domains where the

Table 9.3: Comparison of the $1\sigma$ constraints of cosmological parameters between different methods at fiducial cosmology.

| Dataset | Method | Galaxy Shape Noise | $\Omega_m$ | $\sigma_8$ | $S_8$ |
|---|---|---|---|---|---|
| DM-only maps | ResNet18 + VMIM loss | $n_g = 10\,\mathrm{arcmin}^{-2}$ | 0.076 | 0.114 | 0.046 |
| | | $n_g = 30\,\mathrm{arcmin}^{-2}$ | 0.033 | 0.064 | 0.027 |
| | | $n_g = 100\,\mathrm{arcmin}^{-2}$ | 0.021 | 0.044 | 0.022 |
| | ResNet18 + $\mathrm{MSE_{NP}}$ loss | $n_g = 10\,\mathrm{arcmin}^{-2}$ | 0.091 | 0.126 | 0.047 |
| | | $n_g = 30\,\mathrm{arcmin}^{-2}$ | 0.034 | 0.067 | 0.029 |
| | | $n_g = 100\,\mathrm{arcmin}^{-2}$ | 0.022 | 0.048 | 0.023 |
| | ResNet18 + $\mathrm{MSE_{PCA}}$ | $n_g = 10\,\mathrm{arcmin}^{-2}$ | 0.090 | 0.118 | 0.047 |
| | | $n_g = 30\,\mathrm{arcmin}^{-2}$ | 0.046 | 0.087 | 0.036 |
| | | $n_g = 100\,\mathrm{arcmin}^{-2}$ | 0.023 | 0.045 | 0.023 |
| | ResNet18 + MSE loss | $n_g = 10\,\mathrm{arcmin}^{-2}$ | 0.124 | 0.135 | 0.067 |
| | | $n_g = 30\,\mathrm{arcmin}^{-2}$ | 0.064 | 0.100 | 0.040 |
| | | $n_g = 100\,\mathrm{arcmin}^{-2}$ | 0.024 | 0.046 | 0.024 |
| | Power Spectrum | $n_g = 10\,\mathrm{arcmin}^{-2}$ | 0.124 | 0.160 | 0.061 |
| | | $n_g = 30\,\mathrm{arcmin}^{-2}$ | 0.990 | 0.141 | 0.053 |
| | | $n_g = 100\,\mathrm{arcmin}^{-2}$ | 0.072 | 0.121 | 0.048 |
| Baryon maps | ResNet18 + VMIM loss | $n_g = 20\,\mathrm{arcmin}^{-2}$ | 0.059 | 0.099 | 0.046 |
| | | $n_g = 50\,\mathrm{arcmin}^{-2}$ | 0.033 | 0.071 | 0.038 |
| | | $n_g = 100\,\mathrm{arcmin}^{-2}$ | 0.031 | 0.063 | 0.036 |
| | ResNet18 + $\mathrm{MSE_{NP}}$ loss | $n_g = 20\,\mathrm{arcmin}^{-2}$ | 0.065 | 0.104 | 0.048 |
| | | $n_g = 50\,\mathrm{arcmin}^{-2}$ | 0.043 | 0.078 | 0.039 |
| | | $n_g = 100\,\mathrm{arcmin}^{-2}$ | 0.034 | 0.071 | 0.038 |
| | ResNet18 + $\mathrm{MSE_{PCA}}$ | $n_g = 20\,\mathrm{arcmin}^{-2}$ | 0.064 | 0.106 | 0.049 |
| | | $n_g = 50\,\mathrm{arcmin}^{-2}$ | 0.034 | 0.074 | 0.040 |
| | | $n_g = 100\,\mathrm{arcmin}^{-2}$ | 0.039 | 0.074 | 0.039 |
| | ResNet18 + MSE loss | $n_g = 20\,\mathrm{arcmin}^{-2}$ | 0.069 | 0.111 | 0.050 |
| | | $n_g = 50\,\mathrm{arcmin}^{-2}$ | 0.051 | 0.089 | 0.044 |
| | | $n_g = 100\,\mathrm{arcmin}^{-2}$ | 0.053 | 0.092 | 0.045 |
| | Power Spectrum | $n_g = 20\,\mathrm{arcmin}^{-2}$ | 0.090 | 0.133 | 0.060 |
| | | $n_g = 50\,\mathrm{arcmin}^{-2}$ | 0.065 | 0.107 | 0.052 |
| | | $n_g = 100\,\mathrm{arcmin}^{-2}$ | 0.054 | 0.097 | 0.049 |

Table 9.4: Comparison of the constraining power between different methods in the presence of baryons. Here, we marginalize over baryon parameters.

| Method | $n_g = 20\,\mathrm{arcmin}^{-2}$ | $n_g = 50\,\mathrm{arcmin}^{-2}$ | $n_g = 100\,\mathrm{arcmin}^{-2}$ |
|---|---|---|---|
| ResNet18 + VMIM loss | 112 | 151 | 202 |
| ResNet18 + $\mathrm{MSE_{NP}}$ loss | 96 | 148 | 195 |
| ResNet18 + $\mathrm{MSE_{PCA}}$ loss | 89 | 135 | 179 |
| ResNet18 + MSE loss | 87 | 121 | 150 |
| Power Spectrum | 48 | 68 | 84 |
| [198] CNN | $\sim 77$ | $\sim 109$ | $\sim 136$ |
| Multiscale Flow ($256^2$) | 137 | 242 | 338 |

amount of training and test data is limited, transfer learning has become an essential tool in modern machine learning research [e.g. 72, 201, 154].

Transfer learning (TL) has also shown promise in astrophysics. [344] demonstrated the effectiveness of transfer learning in classifying stellar light curves, enhancing the accuracy of categorizing astrophysical phenomena and understanding of stellar behavior and properties. [332] highlighted solutions that transfer learning introduces, particularly in dealing with the diverse range of data types encountered in astronomy.

We use three different datasets for getting our pre-trained models. First, we harness models pre-trained on the extensively benchmarked ImageNet dataset, known for its comprehensive compilation of image classification challenges. Next, we pre-train our models on gaussian random fields — a cost-effective option due to their straightforward production process. These models are then fine-tuned on our WL maps. Finally, we extend our approach to include models initially trained on log-normal fields, and these models also undergo subsequent refinement on our WL maps.

By leveraging a pre-trained neural network, our CNNs are endowed with a rich set of features and insights garnered from extensive datasets, which are then intricately honed to suit the nuances of our domain-specific task of cosmological inference. The fine-tuning phase is critical; it enables the pre-existing neural architecture to recalibrate, aligning its learned patterns and intricacies with the unique characteristics of our WL maps. Our results of the constraining powers achieved using transfer learning are shown in tables 9.5 and 9.6.

## ImageNet dataset TL

We used publicly available ResNet18 models pre-trained on the ImageNet dataset [122]. The validation accuracy on the ImageNet dataset of these models is $\sim 90\%$. The pre-trained models were then fine-tuned on our DM-only WL dataset, and parameter constraints were obtained using the same methodology as training from scratch.

Despite the high validation accuracy on ImageNet, the final models did not attain a level of performance on our WL dataset that was equivalent to that of models trained from scratch with

Table 9.5: Comparison of the constraining power between different methods using Gaussian random maps transfer learning. The results are similar to the results in table 9.1.

| Method | $n_g = 10\,\mathrm{arcmin}^{-2}$ | $n_g = 30\,\mathrm{arcmin}^{-2}$ | $n_g = 100\,\mathrm{arcmin}^{-2}$ |
|---|---|---|---|
| ResNet18 + VMIM loss | 70 | 221 | 412 |
| ResNet18 + $\mathrm{MSE}_{\mathrm{NP}}$ loss | 74 | 154 | 360 |
| ResNet18 + $\mathrm{MSE}_{\mathrm{PCA}}$ loss | 55 | 142 | 281 |
| ResNet18 + MSE loss | 45 | 128 | 286 |

randomly initialized weights. This is not too surprising because of the stark differences in visual features between the everyday objects of the ImageNet dataset and the complex cosmological and astrophysical patterns present in WL maps.

## Pre-training with Gaussian random fields

Next, we consider pre-training the networks on Gaussian Random Fields (GRFs) and test if it helps reduce the overfitting and improves the constraining power. During the pre-training, we randomly sample cosmological parameters from the prior and generate GRFs on the fly with the same power spectra as the WL maps.

We first directly apply this pre-trained network for inference without any fine-tuning on numerical simulations. Note that this network only serves as a data compressor. As long as we measure the mean and covariance matrix in Equation 9.6 with realistic WL maps, the posterior inference won't be biased even if we train the network on GRFs. In this case, we find comparable constraining power as the power spectrum analysis, which is not surprising since the GRFs only contain the power spectrum information.

After fine-tuning the networks on DM-only WL maps, we present the figure of merit in Table 9.5. We find comparable constraining power compared to training the models from scratch. Pre-training the model with GRFs did not improve our posterior constraints.

## Pre-training with log-normal fields

While GRFs provide a reasonable approximation of WL maps on large scales, they lack non-Gaussian features, and pre-training on these GRFs does not help with the modeling of small-scale non-Gaussian structures. We further explore pre-training with log-normal fields, which are a better approximation to the non-Gaussian WL maps and are widely used for their modeling [311, 356, 45].

The log-normal fields $\kappa_{\mathrm{LN}}$ can be generated efficiently with

$$\kappa_{\mathrm{LN}}(\theta) = e^{\kappa_{\mathrm{GRF}}(\theta)} - \lambda, \tag{9.7}$$

where $\kappa_{\mathrm{GRF}}$ is a GRF, and $\lambda$ is referred to as the shift parameter of the lognormal distribution. The shift parameter depends on the scale at which the field is smoothed or pixelized. In this work, we

Table 9.6: Comparison of the constraining power between different methods using lognormal maps transfer learning. The results are similar to the results in table 9.1.

| Method | $n_g = 10\,\mathrm{arcmin}^{-2}$ | $n_g = 30\,\mathrm{arcmin}^{-2}$ | $n_g = 100\,\mathrm{arcmin}^{-2}$ |
|---|---|---|---|
| ResNet18 + VMIM loss | 65 | 240 | 396 |
| ResNet18 + $\mathrm{MSE_{NP}}$ loss | 67 | 165 | 347 |
| ResNet18 + $\mathrm{MSE_{PCA}}$ loss | 48 | 130 | 295 |
| ResNet18 + MSE loss | 50 | 118 | 300 |

measure $\lambda$ from the mock WL maps. The correlation function of $\kappa_{\rm LN}$ is related to the correlation function of $\kappa_{\rm GRF}$ via [356]:

$$\xi_{\rm GRF}(\theta) = \log\left[\frac{\xi_{LN}(\theta)}{\alpha^2} + 1\right], \tag{9.8}$$

where $\alpha = \exp\left(\mu + \sigma^2/2\right)$, with $\mu$ and $\sigma^2$ being the mean and variance of $\kappa_{\rm GRF}$.

Similar to section 9.4, during pre-training, we randomly sample cosmological parameters and generate log-normal fields on the fly with the same power spectrum as the simulated WL maps. The power spectrum of $\kappa_{\rm GRF}$ is calculated numerically from Equation 9.8. Then we fine-tune the pre-trained network on the simulated WL maps and perform posterior inference following the same procedure described in section 9.2.

In table 9.6 we show the figure of merit of posterior constraints with networks pre-trained on log-normal maps. Again, we find comparable performance as compared to training the network from scratch. This is possibly due to the specific non-Gaussian features inherent in the weak lensing data that are not fully captured by the log-normal approximation.

## 9.5 Conclusions

In this paper, we trained convolutional neural networks to constrain the underlying cosmological parameters ($\sigma_8$, $\Omega_m$) of simulated weak lensing convergence maps (with and without baryonic effects) in the presence of shape noise levels corresponding to ongoing and future large weak lensing surveys.

Our results indicate that discriminative CNN models, when trained using Mutual Information Maximization-based and principal component analysis-enhanced loss functions, yield significantly tighter constraints on the $\Omega_m - \sigma_8$ parameter space than conventional methods such as power spectrum analysis, peak counting, and CNN models trained with MSE loss.

Notably, the Variational Mutual Information Maximization (VMIM) loss function leads to an improvement in parameter constraints, using DM-only WL maps, by factors of $\sim 5$ over the power spectrum, $\sim 3$ over peak counts, and $\sim 2$ over previous CNN models, even under varying noise levels representative of surveys like LSST and Euclid. Using noisy WL convergence maps with

baryons, we achieve $\sim 2.3$ times stronger constraining power than the power spectrum and $\sim 1.5$ over previous CNN models.

To further explore the possibilities of improvement for this task, we also discussed transfer learning where we adapted pre-trained models, also called Foundation Models, trained on a large number of different tasks or datasets, for cosmological inference. However, our results suggest that the direct training on weak lensing maps still holds a slight edge, and pretraining on a large number of Gaussian or log-normal maps provided no advantage. This is possibly due to the specific non-Gaussian features inherent in the weak lensing data that are not fully captured by the log-normal or Gaussian random approximations. This suggests that the use of Foundation Models may be of limited use in some scientific tasks, specially when the Foundation Model training data differ from the real data in some important aspects and there is enough realistic training data available so that there is no need for Transfer Learning.

In previous work generative MultiScale Flow (MSF) [55] outperformed discriminative CNN, but for realistic noise levels such as 10 or 30 $\mathrm{arcmin}^{-2}$ we find that our CNN results are comparable, potentially suggesting both may have extracted the full information content from the training data. For optimistic noise levels like 50 or 100 $\mathrm{arcmin}^{-2}$, MSF outperforms CNN by up to 60% in Figure of Merit. Assuming they both give equal performance the choice of the method will thus depend on other requirements: discriminative CNN is simpler to train, and scales well to the higher dimensional data, while generative Normalizing Flow models also provide additional tests of unknown systematics such as density estimation of the data (a generalized goodness of fit test). We expect that future applications on simulated and real data will continue to explore both of these approaches to extract maximal amount of information from the data, and to assess its robustness against systematics and other effects.

In this paper, to directly compare our results with previous works, we adopt the same simple setup in our analysis as previous studies and do not include realistic effects like source galaxy redshift distribution, tomography, or astrophysical systematics (such as intrinsic alignment, shear multiplication biases, etc.). We plan to explore the impact of these effects on our analysis in future works.

# Acknowledgements

We thank the Columbia Lensing group for making their simulations available and Francois Lanusse for useful discussions and software.

# Chapter 10

# Conclusions

The standard cosmological model ΛCDM has been a successful framework for predicting and explaining various observations. Nevertheless, it faces several challenges and significant components of the model remain mysterious, such as the nature of dark energy, and the origin of $S_8$ tension. Could they lead to new physics? To answer these questions, current and upcoming sky surveys such as HSC [3], DESI [66], Euclid[9], Rubin Observatory [195], and Roman space telescope [75] aim to measure the large-scale structures and evolution of the universe with multiple probes, allowing us to constrain cosmological parameters to high precision and search for deviations from ΛCDM. However, traditional analysis based on two-point statistics can only extract a limited amount of information from these non-Gaussian data. Improved theoretical modeling and data analysis methods are necessary to fully realize the potential of these observations.

Leveraging high-fidelity cosmological simulations, simulation-based inference (SBI) [51] provides a promising framework for extracting rich non-Gaussian information from non-linear scales. However, several challenges limit its cosmological applications. On the simulation side, it is computationally challenging to run a large number of high-resolution large-volume cosmological simulations to fully sample the parameter space for training the SBI models. Reducing the simulation resolution and volume leads to model misspecification and could bias the cosmological constraints. On the inference side, while performing SBI on low-dimensional summary statistics is straightforward, field-level SBI can be challenging due to the high dimensionality of the data.

To facilitate the deployment of SBI approaches into upcoming survey analysis pipelines, in this dissertation, we develop several physics-motivated machine-learning models to improve both the numerical simulations and the SBI models. Fast N-body PM simulations are able to simulate the LSS with low computational costs, making them suitable for the training SBI models, but their lack of small-scale force and baryonic physics and long-time steps cannot accurately predict small-scale matter distribution. In chapter 2, we develop a Potential Gradient Descent (PGD) model to improve the modeling of nonlinear gravity on small scales. We also introduce the Enthalpy Gradient Descent (EGD) model to effectively model the baryonic feedback. We show that these models are able to improve the small-scale matter power spectrum, the cross-correlation coefficient with full N-body simulation or hydrodynamical simulations, the halo profiles, and the identification of subhalos. In chapter 3, we incorporate this model into lightcone simulations, and show that it

significantly improves the small-scale modeling of weak gravitational lensing. In chapter 4, we further develop a public differentiable package MADlens for accurately generating weak lensing predictions with low computational cost. These developments focus on improving the small-scale matter distribution, primarily targeting weak lensing analysis. In terms of galaxy clustering, we develop a relaxed-FOF halo finder in chapter 3, for improving the halo statistics from low-resolution fast N-body simulations.

In chapter 5 and 6, we focus on the modeling of baryons in N-body simulations. We develop a Lagrangian Deep Learning (LDL) model for modeling various baryonic observables in chapter 5. We combine the N-body solver FastPM with LDL and apply them to a wide range of cosmological outputs, from dark matter to stellar maps, gas density, and temperature. The computational cost of LDL is nearly four orders of magnitude lower than the full hydrodynamical simulations, yet it outperforms hydro simulations at the same resolution. This opens up the possibility of analyzing cosmological observations entirely within this framework, without the need for large full N-body simulations. In chapter 6, we focus on modeling the baryonic effect on the total matter distribution for weak lensing applications. By analyzing thousands of hydrodynamical simulations, we find an efficient way to model the baryonic feedback and build an emulator at the field level using Gaussian processes. This method and our emulator enable field-level SBI analyses and accounting for baryonic effects in weak lensing analyses.

In the second half of this dissertation (chapter 7, 8 and 9), we focus on the development of SBI models. Off-the-shelf ML models are not curated for astrophysical applications, and thus do not take advantage of the characteristics of cosmological data (e.g., symmetries) or satisfy scientific needs (i.e., well-calibrated and robust parameter inference). We integrate physics constraints and domain knowledge into ML models for superior performance in scientific applications. In chapter 7, we develop a Translation and Rotation Equivariant Normalizing Flow (TRENF) model that satisfies symmetry constraints. We apply TRENF to learn the high-dimensional field-level data likelihood, and show that it leads to significant improvements in constraining power over the standard power spectrum summary statistic. In chapter 8, we introduce Multiscale Flow (MSF) for optimal and robust field-level analysis. MSF uses hierarchical decomposition of cosmological fields via a wavelet basis, and then models different wavelet components separately as Normalizing Flows. This decomposition allows us to separate the information from different scales and identify distribution shifts in the data such as unknown scale-dependent systematics. We show that the resulting likelihood analysis can not only identify these types of systematics (such as baryonic effect), but can also be made optimal, in the sense that the Multiscale Flow can learn the full likelihood at the field without any dimensionality reduction. Finally, the SBI machine learning models usually come with a large number of hyperparameters and training choices, making them hard to use in practice. In chapter 9, we perform a comparative study and explore the impact of different analysis choices on parameter constraints for a Convolutional Neural Networks (CNN) based SBI model. By performing a large number of control experiments, we identify the best hyperparameter choice and training loss functions that improve on previous CNN-based SBI models by a factor of 2.

These developments mark a substantial step toward the full deployment of SBI approaches into cosmological survey analysis pipelines, offering the promise of a deeper understanding of our Universe and the potential discovery of new physics beyond the current model.

# Bibliography


[1] Kevork N. Abazajian et al. "CMB-S4 Science Book, First Edition". In: *arXiv e-prints*, arXiv:1610.02743 (Oct. 2016), arXiv:1610.02743. DOI: 10.48550/arXiv.1610.02743. arXiv: 1610.02743 [astro-ph.CO].

[2] T. M. C. Abbott et al. "Dark Energy Survey Year 3 results: Cosmological constraints from galaxy clustering and weak lensing". In: *Phys. Rev. D* 105.2, 023520 (Jan. 2022), p. 023520. DOI: 10.1103/PhysRevD.105.023520. arXiv: 2105.13549 [astro-ph.CO].

[3] Hiroaki Aihara et al. "The Hyper Suprime-Cam SSP Survey: Overview and survey design". In: *PASJ* 70, S4 (Jan. 2018), S4. DOI: 10.1093/pasj/psx066. arXiv: 1704.05858 [astro-ph.IM].

[4] Hiroaki Aihara et al. "The Hyper Suprime-Cam SSP survey: overview and survey design". In: *Publications of the Astronomical Society of Japan* 70.SP1 (2018), S4.

[5] Erwan Allys et al. "New interpretable statistics for large-scale structure analysis and generation". In: *Physical Review D* 102.10 (2020), p. 103506.

[6] Justin Alsing, Alan Heavens, and Andrew H. Jaffe et al. "Hierarchical cosmic shear power spectrum inference". In: *MNRAS* 455.4 (Feb. 2016), pp. 4452–4466. DOI: 10.1093/mnras/stv2501. arXiv: 1505.07840 [astro-ph.CO].

[7] Justin Alsing, Benjamin Wandelt, and Stephen Feeney. "Massive optimal data compression and density estimation for scalable, likelihood-free inference in cosmology". In: *Monthly Notices of the Royal Astronomical Society* 477.3 (2018), pp. 2874–2885.

[8] Justin Alsing et al. "Fast likelihood-free cosmology with neural density estimators and active learning". In: *Monthly Notices of the Royal Astronomical Society* 488.3 (2019), pp. 4440–4458.

[9] Luca Amendola et al. "Cosmology and fundamental physics with the Euclid satellite". In: *Living Reviews in Relativity* 21.1, 2 (Apr. 2018), p. 2. DOI: 10.1007/s41114-017-0010-3. arXiv: 1606.00180 [astro-ph.CO].

[10] Raul E Angulo et al. "The BACCO simulation project: exploiting the full power of large-scale structure for cosmology". In: *Monthly Notices of the Royal Astronomical Society* 507.4 (July 2021), pp. 5869–5881. ISSN: 1365-2966. DOI: 10.1093/mnras/stab2018. URL: http://dx.doi.org/10.1093/mnras/stab2018.

[11] Giovanni Aricò et al. "Modelling the large-scale mass density field of the universe as a function of cosmology and baryonic physics". In: *Monthly Notices of the Royal Astronomical Society* 495.4 (July 2020), pp. 4800–4819. DOI: 10.1093/mnras/staa1478. arXiv: 1911.08471 [astro-ph.CO].

[12] Giovanni Aricò et al. "The BACCO simulation project: a baryonification emulator with neural networks". In: *MNRAS* 506.3 (Sept. 2021), pp. 4070–4082. DOI: 10.1093/mnras/stab1911. arXiv: 2011.15018 [astro-ph.CO].

[13] J. S. Bagla. "TreePM: A Code for Cosmological N-Body Simulations". In: *Journal of Astrophysics and Astronomy* 23.3-4 (Dec. 2002), pp. 185–196. DOI: 10.1007/BF02702282. arXiv: astro-ph/9911025 [astro-ph].

[14] David Barber and Felix Agakov. "The IM Algorithm: A Variational Approach to Information Maximization." In: Jan. 2003.

[15] Josh Barnes and Piet Hut. "A hierarchical O(N log N) force-calculation algorithm". In: *Nature* 324.6096 (Dec. 1986), pp. 446–449. DOI: 10.1038/324446a0.

[16] Matthias Bartelmann and Matteo Maturi. "Weak gravitational lensing". In: *Scholarpedia* 12.1 (Jan. 2017), p. 32440. DOI: 10.4249/scholarpedia.32440. arXiv: 1612.06535 [astro-ph.CO].

[17] Matthias Bartelmann and Peter Schneider. "Weak gravitational lensing". In: *Physics Reports* 340.4-5 (2001), pp. 291–472.

[18] Peter Behroozi et al. "UniverseMachine: The correlation between galaxy growth and dark matter halo assembly from z= 0- 10". In: *Monthly Notices of the Royal Astronomical Society* 488.3 (2019), pp. 3143–3194.

[19] Peter S. Behroozi, Risa H. Wechsler, and Hao-Yi Wu. "The ROCKSTAR Phase-space Temporal Halo Finder and the Velocity Offsets of Cluster Cores". In: *ApJ* 762.2, 109 (Jan. 2013), p. 109. DOI: 10.1088/0004-637X/762/2/109. arXiv: 1110.4372 [astro-ph.CO].

[20] Philippe Berger and George Stein. "A volumetric deep convolutional neural network for simulation of mock dark matter halo catalogues". In: *Monthly Notices of the Royal Astronomical Society* 482.3 (2019), pp. 2861–2871.

[21] Andreas A. Berlind and David H. Weinberg. "The Halo Occupation Distribution: Toward an Empirical Determination of the Relation between Galaxies and Mass". In: *ApJ* 575.2 (Aug. 2002), pp. 587–616. DOI: 10.1086/341469. arXiv: astro-ph/0109001 [astro-ph].

[22] Mauro Bernardini et al. "Predicting dark matter halo formation in N-body simulations with deep regression networks". In: *Monthly Notices of the Royal Astronomical Society* 496.4 (2020), pp. 5116–5125.

[23] Simeon Bird et al. "An emulator for the Lyman-$\alpha$ forest". In: *Journal of Cosmology and Astroparticle Physics* 2019.02 (Feb. 2019), pp. 050–050. ISSN: 1475-7516. DOI: 10.1088/1475-7516/2019/02/050. URL: http://dx.doi.org/10.1088/1475-7516/2019/02/050.

[24] Simeon Bird et al. "The ASTRID simulation: galaxy formation and reionization". In: *MNRAS* 512.3 (May 2022), pp. 3703–3716. DOI: 10.1093/mnras/stac648. arXiv: 2111.01160 [astro-ph.GA].

[25] Michael R. Blanton et al. "Sloan Digital Sky Survey IV: Mapping the Milky Way, Nearby Galaxies, and the Distant Universe". In: *AJ* 154.1, 28 (July 2017), p. 28. DOI: 10.3847/1538-3881/aa7567. arXiv: 1703.00052 [astro-ph.GA].

[26] Diego Blas, Julien Lesgourgues, and Thomas Tram. "The Cosmic Linear Anisotropy Solving System (CLASS). Part II: Approximation schemes". In: *J. Cosmology Astropart. Phys.* 2011.7, 034 (July 2011), p. 034. DOI: 10.1088/1475-7516/2011/07/034. arXiv: 1104.2933 [astro-ph.CO].

[27] Sebastian Bocquet et al. "The Mira-Titan Universe. III. Emulation of the Halo Mass Function". In: *ApJ* 901.1, 5 (Sept. 2020), p. 5. DOI: 10.3847/1538-4357/abac5c. arXiv: 2003.12116 [astro-ph.CO].

[28] V. Böhm et al. "MADLens, a python package for fast and differentiable non-Gaussian lensing simulations". In: *Astronomy and Computing* 36, 100490 (July 2021), p. 100490. DOI: 10.1016/j.ascom.2021.100490. arXiv: 2012.07266 [astro-ph.CO].

[29] Vanessa Böhm, Stefan Hilbert, and Maksim Greiner et al. "Bayesian weak lensing tomography: Reconstructing the 3D large-scale distribution of matter with a lognormal prior". In: *Phys. Rev. D* 96.12, 123510 (Dec. 2017), p. 123510. DOI: 10.1103/PhysRevD.96.123510. arXiv: 1701.01886 [astro-ph.CO].

[30] J. R. Bond, A. H. Jaffe, and L. Knox. "Estimating the power spectrum of the cosmic microwave background". In: *Physical Review D* 57.4 (Feb. 1998), pp. 2117–2137. DOI: 10.1103/PhysRevD.57.2117. arXiv: astro-ph/9708203 [astro-ph].

[31] J. R. Bond and A. S. Szalay. "The collisionless damping of density fluctuations in an expanding universe". In: *ApJ* 274 (Nov. 1983), pp. 443–468. DOI: 10.1086/161460.

[32] Guillaume Bouchard. "Bias-variance tradeoff in hybrid generative-discriminative models". In: *Sixth International Conference on Machine Learning and Applications (ICMLA 2007)*. IEEE. 2007, pp. 124–129.

[33] Stevo Bozinovski. "Reminder of the first paper on transfer learning in neural networks, 1976". In: *Informatica* 44.3 (2020).

[34] Stevo Bozinovski and Ante Fulgosi. "The influence of pattern similarity and transfer learning upon training of a base perceptron b2". In: *Proceedings of Symposium Informatica*. Vol. 3. 1976, pp. 121–126.

[35] James Bradbury et al. "JAX: composable transformations of Python+ NumPy programs". In: (2018).

[36] Richard H Byrd et al. "A limited memory algorithm for bound constrained optimization". In: *SIAM Journal on scientific computing* 16.5 (1995), pp. 1190–1208.

[37] Patrick Cannon, Daniel Ward, and Sebastian M. Schmon. "Investigating the Impact of Model Misspecification in Neural Simulation-based Inference". In: *arXiv e-prints*, arXiv: 2209.01845 (Sept. 2022), arXiv: 2209.01845. DOI: 10.48550/arXiv.2209.01845. arXiv: 2209.01845 [stat.ML].

[38] V. F. Cardone, S. Camera, and R. Mainini et al. "Weak lensing peak count as a probe of f(R) theories". In: *MNRAS* 430.4 (Apr. 2013), pp. 2896–2909. DOI: 10.1093/mnras/stt084. arXiv: 1204.3148 [astro-ph.CO].

[39] John Joseph M Carrasco, Mark P Hertzberg, and Leonardo Senatore. "The effective field theory of cosmological large scale structures". In: *Journal of High Energy Physics* 2012.9 (2012), p. 82.

[40] Pedro Carrilho, Chiara Moretti, and Alkistis Pourtsidou. "Cosmology with the EFTofLSS and BOSS: dark energy constraints and a note on priors". In: *J. Cosmology Astropart. Phys.* 2023.1, 028 (Jan. 2023), p. 028. DOI: 10.1088/1475-7516/2023/01/028. arXiv: 2207.14784 [astro-ph.CO].

[41] Tom Charnock, Guilhem Lavaux, and Benjamin D. Wandelt. "Automatic physical inference with information maximizing neural networks". In: *Physical Review D* 97.8, 083004 (Apr. 2018), p. 083004. DOI: 10.1103/PhysRevD.97.083004. arXiv: 1802.03537 [astro-ph.IM].

[42] Tom Charnock et al. "Neural physical engines for inferring the halo mass distribution function". In: *Monthly Notices of the Royal Astronomical Society* 494.1 (May 2020), pp. 50–61. DOI: 10.1093/mnras/staa682. arXiv: 1909.06379 [astro-ph.CO].

[43] Sihao Cheng and Brice Ménard. "Weak lensing scattering transform: dark energy and neutrino mass sensitivity". In: *Monthly Notices of the Royal Astronomical Society* 507.1 (July 2021), pp. 1012–1020. ISSN: 1365-2966. DOI: 10.1093/mnras/stab2102. URL: http://dx.doi.org/10.1093/mnras/stab2102.

[44] Sihao Cheng et al. "A new approach to observational cosmology using the scattering transform". In: *Monthly Notices of the Royal Astronomical Society* 499.4 (Dec. 2020), pp. 5902–5914. DOI: 10.1093/mnras/staa3165. arXiv: 2006.08561 [astro-ph.CO].

[45] Lucinda Clerkin et al. "Testing the lognormality of the galaxy and weak lensing convergence distributions from Dark Energy Survey maps". In: *Monthly Notices of the Royal Astronomical Society* 466.2 (2017), pp. 1444–1461.

[46] Taco Cohen and Max Welling. "Group equivariant convolutional networks". In: *International conference on machine learning*. PMLR. 2016, pp. 2990–2999.

[47] Taco S. Cohen, Mario Geiger, and Maurice Weiler. "A General Theory of Equivariant CNNs on Homogeneous Spaces". In: *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*. Ed. by Hanna M. Wallach et al. 2019, pp. 9142–9153. URL: https://proceedings.neurips.cc/paper/2019/hash/b9cfe8b6042cf759dc4c0cccb27a6737-Abstract.html.

[48] Taco S. Cohen and Max Welling. "Steerable CNNs". In: *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net, 2017. URL: https://openreview.net/forum?id=rJQKYt5ll.

[49] William R. Coulton, Jia Liu, and Mathew S. Madhavacheril et al. "Constraining neutrino mass with the tomographic weak lensing bispectrum". In: *J. Cosmology Astropart. Phys.* 2019.5, 043 (May 2019), p. 043. DOI: 10.1088/1475-7516/2019/05/043. arXiv: 1810.02374 [astro-ph.CO].

[50] Robert A. Crain et al. "The EAGLE simulations of galaxy formation: calibration of subgrid physics and model variations". In: *MNRAS* 450.2 (June 2015), pp. 1937–1961. DOI: 10.1093/mnras/stv725. arXiv: 1501.01311 [astro-ph.GA].

[51] Kyle Cranmer, Johann Brehmer, and Gilles Louppe. "The frontier of simulation-based inference". In: *Proceedings of the National Academy of Sciences* 117.48 (2020), pp. 30055–30062.

[52] Carolina Cuesta-Lazaro and Siddharth Mishra-Sharma. "Diffusion generative modeling for galaxy surveys: emulating clustering for inference at the field level". In: (2023).

[53] Biwei Dai, Yu Feng, and Uroš Seljak. "A gradient based method for modeling baryons and matter in halos of fast simulations". In: *J. Cosmology Astropart. Phys.* 2018.11, 009 (Nov. 2018), p. 009. DOI: 10.1088/1475-7516/2018/11/009. arXiv: 1804.00671 [astro-ph.CO].

[54] Biwei Dai and Uroš Seljak. "Learning effective physical laws for generating cosmological hydrodynamics with Lagrangian deep learning". In: *Proceedings of the National Academy of Science* 118.16, e2020324118 (Apr. 2021), e2020324118. DOI: 10.1073/pnas.2020324118. arXiv: 2010.02926 [astro-ph.CO].

[55] Biwei Dai and Uroš Seljak. "Multiscale Flow for robust and optimal cosmological analysis". In: *Proceedings of the National Academy of Science* 121.9, e2309624121 (Feb. 2024), e2309624121. DOI: 10.1073/pnas.2309624121.

[56] Biwei Dai and Uroš Seljak. "Translation and rotation equivariant normalizing flow (TRENF) for optimal cosmological analysis". In: *MNRAS* 516.2 (Oct. 2022), pp. 2363–2373. DOI: 10.1093/mnras/stac2010. arXiv: 2202.05282 [astro-ph.CO].

[57] Biwei Dai and Uros Seljak. "Sliced Iterative Normalizing Flows". In: *Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event*. Ed. by Marina Meila and Tong Zhang. Vol. 139. Proceedings of Machine Learning Research. PMLR, 2021, pp. 2352–2364. URL: http://proceedings.mlr.press/v139/dai21a.html.

[58] Biwei Dai et al. "High mass and halo resolution from fast low resolution simulations". In: *J. Cosmology Astropart. Phys.* 2020.4, 002 (Apr. 2020), p. 002. DOI: 10.1088/1475-7516/2020/04/002. arXiv: 1908.05276 [astro-ph.CO].

[59] Ingrid Daubechies. "Orthonormal bases of compactly supported wavelets". In: *Communications on pure and applied mathematics* 41.7 (1988), pp. 909–996.

[60] Romeel Davé, Robert Thompson, and Philip F. Hopkins. "MUFASA: galaxy formation simulations with meshless hydrodynamics". In: *MNRAS* 462.3 (Nov. 2016), pp. 3265–3284. DOI: 10.1093/mnras/stw1862. arXiv: 1604.01418 [astro-ph.GA].

[61] Romeel Davé et al. "SIMBA: Cosmological simulations with black hole growth and feedback". In: *Monthly Notices of the Royal Astronomical Society* 486.2 (2019), pp. 2827–2849.

[62] J. DeRose et al. "The AEMULUS Project. I. Numerical Simulations for Precision Cosmology". In: *ApJ* 875, 69 (Apr. 2019), p. 69. DOI: 10.3847/1538-4357/ab1085. arXiv: 1804.05865.

[63] Joseph DeRose et al. "The Aemulus Project. I. Numerical Simulations for Precision Cosmology". In: *The Astrophysical Journal* 875.1 (2019), p. 69.

[64] DES Collaboration et al. "Dark Energy Survey Year 1 results: Cosmological constraints from cosmic shear". In: *Phys. Rev. D* 98.4, 043528 (Aug. 2018), p. 043528. DOI: 10.1103/PhysRevD.98.043528. arXiv: 1708.01538 [astro-ph.CO].

[65] DESI Collaboration et al. "DESI 2024 VI: Cosmological Constraints from the Measurements of Baryon Acoustic Oscillations". In: *arXiv e-prints*, arXiv:2404.03002 (Apr. 2024), arXiv:2404.03002. DOI: 10.48550/arXiv.2404.03002. arXiv: 2404.03002 [astro-ph.CO].

[66] DESI Collaboration et al. "The DESI Experiment Part I: Science,Targeting, and Survey Design". In: *arXiv e-prints*, arXiv:1611.00036 (Oct. 2016), arXiv:1611.00036. DOI: 10.48550/arXiv.1611.00036. arXiv: 1611.00036 [astro-ph.IM].

[67] Tiziana Di Matteo, Volker Springel, and Lars Hernquist. "Energy input from quasars regulates the growth and activity of black holes and their host galaxies". In: *Nature* 433.7026 (Feb. 2005), pp. 604–607. ISSN: 1476-4687. DOI: 10.1038/nature03335. URL: http://dx.doi.org/10.1038/nature03335.

[68] Jürg Diemand, Michael Kuhlen, and Piero Madau. "Early Supersymmetric Cold Dark Matter Substructure". In: *ApJ* 649.1 (Sept. 2006), pp. 1–13. DOI: 10.1086/506377. arXiv: astro-ph/0603250 [astro-ph].

[69] J. P. Dietrich and J. Hartlap. "Cosmology with the shear-peak statistics". In: *MNRAS* 402.2 (Feb. 2010), pp. 1049–1058. DOI: 10.1111/j.1365-2966.2009.15948.x. arXiv: 0906.3512 [astro-ph.CO].

[70] Laurent Dinh, David Krueger, and Yoshua Bengio. "Nice: Non-linear independent components estimation". In: *arXiv preprint arXiv:1410.8516* (2014).

[71] Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. "Density estimation using Real NVP". In: *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net, 2017. URL: https://openreview.net/forum?id=HkpbnH9lx.

[72] Chuong B Do and Andrew Y Ng. "Transfer learning for text classification". In: *Advances in neural information processing systems* 18 (2005).

[73] Scott Dodelson. *Modern Cosmology*. Amsterdam: Academic Press, 2003. ISBN: 978-0-12-219141-1.

[74] Olivier Doré et al. "Cosmology with the SPHEREX All-Sky Spectral Survey". In: *arXiv e-prints*, arXiv:1412.4872 (Dec. 2014), arXiv:1412.4872. DOI: 10.48550/arXiv.1412.4872. arXiv: 1412.4872 [astro-ph.CO].

[75] Olivier Doré et al. "WFIRST Science Investigation Team "Cosmology with the High Latitude Survey" Annual Report 2017". In: *arXiv e-prints*, arXiv: 1804.03628 (Apr. 2018), arXiv: 1804.03628. DOI: 10.48550/arXiv.1804.03628. arXiv: 1804.03628 [astro-ph.CO].

[76] Cyrille Doux et al. "Dark energy survey internal consistency tests of the joint cosmological probes analysis with posterior predictive distributions". In: *Monthly Notices of the Royal Astronomical Society* 503.2 (2021), pp. 2688–2705.

[77] Simon Duane et al. "Hybrid monte carlo". In: *Physics letters B* 195.2 (1987), pp. 216–222.

[78] Y. Dubois et al. "Dancing in the dark: galactic properties trace spin swings along the cosmic web". In: *MNRAS* 444.2 (Oct. 2014), pp. 1453–1468. DOI: 10.1093/mnras/stu1227. arXiv: 1402.1165 [astro-ph.CO].

[79] A. R. Duffy et al. "Impact of baryon physics on dark matter structures: a detailed simulation study of halo density profiles". In: *MNRAS* 405 (July 2010), pp. 2161–2178. DOI: 10.1111/j.1365-2966.2010.16613.x. arXiv: 1001.3447 [astro-ph.CO].

[80] Conor Durkan et al. "Neural Spline Flows". In: *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*. Ed. by Hanna M. Wallach et al. 2019, pp. 7509–7520. URL: https://proceedings.neurips.cc/paper/2019/hash/7ac71d433f282034e088473244df8c02-Abstract.html.

[81] Daniel J Eisenstein et al. "SDSS-III: Massive spectroscopic surveys of the distant universe, the Milky Way, and extra-solar planetary systems". In: *The Astronomical Journal* 142.3 (2011), p. 72.

[82] Daniel J. Eisenstein and Wayne Hu. "Power Spectra for Cold Dark Matter and Its Variants". In: *ApJ* 511.1 (Jan. 1999), pp. 5–15. DOI: 10.1086/306640. arXiv: astro-ph/9710252 [astro-ph].

[83] Pascal J Elahi et al. "nIFTY galaxy cluster simulations–III. The similarity and diversity of galaxies and subhaloes". In: *Monthly Notices of the Royal Astronomical Society* 458.1 (2016), pp. 1096–1116.

[84] Euclid Collaboration et al. "Euclid preparation: II. The EUCLIDEMULATOR - a tool to compute the cosmology dependence of the nonlinear matter power spectrum". In: *MNRAS* 484.4 (Apr. 2019), pp. 5509–5529. DOI: 10.1093/mnras/stz197. arXiv: 1809.04695 [astro-ph.CO].

[85] Euclid Collaboration et al. "Euclid preparation. VI. Verifying the performance of cosmic shear experiments". In: *A&A* 635, A139 (Mar. 2020), A139. DOI: 10.1051/0004-6361/201936980. arXiv: 1910.10521 [astro-ph.CO].

[86] Euclid Collaboration et al. "Euclid preparation. XXVIII. Forecasts for ten different higher-order weak lensing statistics". In: *A&A* 675, A120 (July 2023), A120. DOI: 10.1051/0004-6361/202346017. arXiv: 2301.12890 [astro-ph.CO].

[87] A. E. Evrard et al. "Virial Scaling of Massive Dark Matter Halos: Why Clusters Prefer a High Normalization Cosmology". In: *ApJ* 672 (Jan. 2008), pp. 122–137. DOI: 10.1086/521616. eprint: astro-ph/0702241.

[88] Yu Feng. *rainwoodman/fastpm-python 0.0.6*. Nov. 2017. DOI: 10.5281/zenodo.1051310. URL: https://doi.org/10.5281/zenodo.1051310.

[89] Yu Feng, Man-Yat Chu, and Uroš Seljak et al. "FASTPM: a new scheme for fast simulations of dark matter and haloes". In: *MNRAS* 463.3 (Dec. 2016), pp. 2273–2286. DOI: 10.1093/mnras/stw2123. arXiv: 1603.00476 [astro-ph.CO].

[90] Yu Feng, Nick Hand, and biweidai. *rainwoodman/kdcount 0.3.27*. Nov. 2017. DOI: 10.5281/zenodo.1051242. URL: https://doi.org/10.5281/zenodo.1051242.

[91] Yu Feng et al. *MP-Gadget/MP-Gadget: A tag for getting a DOI*. Version FirstDOI. Oct. 2018. DOI: 10.5281/zenodo.1451799. URL: https://doi.org/10.5281/zenodo.1451799.

[92] Janis Fluri, Tomasz Kacprzak, and Aurelien Lucchi et al. "Cosmological constraints with deep learning from KiDS-450 weak lensing maps". In: *Phys. Rev. D* 100.6, 063514 (Sept. 2019), p. 063514. DOI: 10.1103/PhysRevD.100.063514. arXiv: 1906.03156 [astro-ph.CO].

[93] Janis Fluri et al. "Cosmological constraints from noisy convergence maps through deep learning". In: *Physical Review D* 98.12 (2018), p. 123518.

[94] Janis Fluri et al. "Full w CDM analysis of KiDS-1000 weak lensing maps using deep learning". In: *Physical Review D* 105.8 (2022), p. 083518.

[95] D. Foreman-Mackey et al. "emcee: The MCMC Hammer". In: *PASP* 125 (Mar. 2013), p. 306. DOI: 10.1086/670067. arXiv: 1202.3665 [astro-ph.IM].

[96] Dan Foreman-Mackey et al. *dfm/tinygp: v0.2.3*. Version v0.2.3. Oct. 2022. DOI: 10.5281/zenodo.7269074. URL: https://doi.org/10.5281/zenodo.7269074.

[97] Roy Friedman and Sultan Hassan. "HIGlow: Conditional Normalizing Flows for High-Fidelity HI Map Modeling". In: *arXiv preprint arXiv:2211.12724* (2022).

[98] Liping Fu, Martin Kilbinger, and Thomas Erben et al. "CFHTLenS: cosmological constraints from a combination of cosmic shear two-point and three-point correlations". In: *MNRAS* 441.3 (July 2014), pp. 2725–2743. DOI: 10.1093/mnras/stu754. arXiv: 1404.5469 [astro-ph.CO].

[99] Lehman H. Garrison et al. "Improving initial conditions for cosmological N-body simulations". In: *MNRAS* 461.4 (Oct. 2016), pp. 4125–4145. DOI: 10.1093/mnras/stw1594. arXiv: 1605.02333 [astro-ph.CO].

[100] M. Gatti 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". In: *arXiv e-prints*, arXiv: 2405.10881 (May 2024), arXiv: 2405.10881. DOI: 10.48550/arXiv.2405.10881. arXiv: 2405.10881 [astro-ph.CO].

[101] Matthew Gebhardt et al. "Cosmological baryon spread and impact on matter clustering in CAMELS". In: *MNRAS* 529.4 (Apr. 2024), pp. 4896–4913. DOI: 10.1093/mnras/stae817. arXiv: 2307.11832 [astro-ph.GA].

[102] S. Genel et al. "Introducing the Illustris project: the evolution of galaxy populations across cosmic time". In: *MNRAS* 445 (Nov. 2014), pp. 175–200. DOI: 10.1093/mnras/stu1654. arXiv: 1405.3749.

[103] Mathieu Germain et al. "Made: Masked autoencoder for distribution estimation". In: *International Conference on Machine Learning*. PMLR. 2015, pp. 881–889.

[104] Sambit K. Giri and Aurel Schneider. *BCemu: Model baryonic effects in cosmological simulations*. Astrophysics Source Code Library, record ascl:2308.010. Aug. 2023. ascl: 2308.010.

[105] Elena Giusarma et al. "Learning neutrino effects in Cosmology with Convolutional Neural Networks". In: *arXiv preprint arXiv:1910.04255* 2019 (2019).

[106] N. Y. Gnedin and L. Hui. "Probing the Universe with the Lyalpha forest - I. Hydrodynamics of the low-density intergalactic medium". In: *MNRAS* 296 (May 1998), pp. 44–55. DOI: 10.1046/j.1365-8711.1998.01249.x. eprint: astro-ph/9706219.

[107] Ian J. Goodfellow et al. "Generative Adversarial Nets". In: *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada*. Ed. by Zoubin Ghahramani et al. 2014, pp. 2672–2680.

[108] Krzysztof M. Gorski. "On Determining the Spectrum of Primordial Inhomogeneity from the COBE DMR Sky Maps: Method". In: *The Astrophysical Journal Letters* 430 (Aug. 1994), p. L85. DOI: 10.1086/187444. arXiv: astro-ph/9403066 [astro-ph].

[109] K. M. Górski et al. "HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere". In: *ApJ* 622 (Apr. 2005), pp. 759–771. DOI: 10.1086/427976. eprint: arXiv:astro-ph/0409513.

[110] JA Gregory and R Delbourgo. "Piecewise rational quadratic interpolation to monotonic data". In: *IMA Journal of Numerical Analysis* 2.2 (1982), pp. 123–130.

[111] Antonio CC Guimaraes. "Cosmological model differentiation through weak gravitational lensing". In: *Monthly Notices of the Royal Astronomical Society* 337.2 (2002), pp. 631–640.

[112] Arushi Gupta et al. "Non-Gaussian information from weak lensing data via deep learning". In: *Physical Review D* 97.10, 103515 (May 2018), p. 103515. DOI: 10.1103/PhysRevD.97.103515. arXiv: 1802.01212 [astro-ph.CO].

[113] Alfred Haar. "Zur Theorie der orthogonalen Funktionensysteme". In: *Mathematische Annalen.* 69.3 (1910), pp. 331–371. ISSN: 0025-5831.

[114] ChangHoon Hahn et al. "SIMBIG: The First Cosmological Constraints from Non-Gaussian and Non-Linear Galaxy Clustering". In: *arXiv e-prints*, arXiv: 2310.15246 (Oct. 2023). DOI: 10.48550/arXiv.2310.15246. arXiv: 2310.15246 [astro-ph.CO].

[115] M. Haider et al. "Large-scale mass distribution in the Illustris simulation". In: *MNRAS* 457 (Apr. 2016), pp. 3024–3035. DOI: 10.1093/mnras/stw077. arXiv: 1508.01525.

[116] Takashi Hamana et al. "Cosmological constraints from cosmic shear two-point correlation functions with HSC survey first-year data". In: *Publications of the Astronomical Society of Japan* 72.1 (2020), p. 16.

[117] A. J. S. Hamilton. "Towards optimal measurement of power spectra - I. Minimum variance pair weighting and the Fisher matrix". In: *Monthly Notices of the Royal Astronomical Society* 289.2 (Aug. 1997), pp. 285–294. DOI: 10.1093/mnras/289.2.285. arXiv: astro-ph/9701008 [astro-ph].

[118] Nick Hand et al. "nbodykit: An Open-source, Massively Parallel Toolkit for Large-scale Structure". In: *ApJ* 156.4, 160 (Oct. 2018), p. 160. DOI: 10.3847/1538-3881/aadae0. arXiv: 1712.05834 [astro-ph.IM].

[119] J. Harnois-Déraps, D. Munshi, and P. Valageas et al. "Testing modified gravity with cosmic shear". In: *MNRAS* 454.3 (Dec. 2015), pp. 2722–2735. DOI: 10.1093/mnras/stv2120. arXiv: 1506.06313 [astro-ph.CO].

[120] Joachim Harnois-Déraps et al. "Cosmic shear cosmology beyond two-point statistics: a combined peak count and correlation function analysis of DES-Y1". In: *Monthly Notices of the Royal Astronomical Society* 506.2 (2021), pp. 1623–1650.

[121] Sultan Hassan et al. "HIFLOW: Generating Diverse HI Maps and Inferring Cosmology while Marginalizing over Astrophysics Using Normalizing Flows". In: *ApJ* 937.2, 83 (Oct. 2022), p. 83. DOI: 10.3847/1538-4357/ac8b09. arXiv: 2110.02983 [astro-ph.CO].

[122] K. He et al. "Deep Residual Learning for Image Recognition". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Los Alamitos, CA, USA: IEEE Computer Society, June 2016, pp. 770–778. DOI: 10.1109/CVPR.2016.90. URL: https://doi.ieeecomputersociety.org/10.1109/CVPR.2016.90.

[123] Kaiming He et al. "Identity mappings in deep residual networks". In: *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14*. Springer. 2016, pp. 630–645.

[124] Siyu He et al. "Learning to predict the cosmological structure formation". In: *Proceedings of the National Academy of Sciences* 116.28 (2019), pp. 13825–13832.

[125] A. F. Heavens, T. D. Kitching, and L. Verde. "On model selection forecasting, dark energy and modified gravity". In: *MNRAS* 380.3 (Sept. 2007), pp. 1029–1035. DOI: 10.1111/j.1365-2966.2007.12134.x. arXiv: astro-ph/0703191 [astro-ph].

[126] Katrin Heitmann et al. "The Coyote Universe Extended: Precision Emulation of the Matter Power Spectrum". In: *ApJ* 780.1, 111 (Jan. 2014), p. 111. DOI: 10.1088/0004-637X/780/1/111. arXiv: 1304.7849 [astro-ph.CO].

[127] Wojciech A. Hellwing et al. "The effect of baryons on redshift space distortions and cosmic density and velocity fields in the EAGLE simulation". In: *Monthly Notices of the Royal Astronomical Society: Letters* 461.1 (Apr. 2016), pp. L11–L15. ISSN: 1745-3933. DOI: 10.1093/mnrasl/slw081. URL: http://dx.doi.org/10.1093/mnrasl/slw081.

[128] Joeri Hermans et al. "A trust crisis in simulation-based inference? your posterior approximations can be unfaithful". In: *arXiv preprint arXiv:2110.06581* (2021).

[129] Catherine Heymans, Emma Grocutt, and Alan Heavens et al. "CFHTLenS tomographic weak lensing cosmological parameter constraints: Mitigating the impact of intrinsic galaxy alignments". In: *MNRAS* 432.3 (July 2013), pp. 2433–2453. DOI: 10.1093/mnras/stt601. arXiv: 1303.1808 [astro-ph.CO].

[130] Catherine Heymans et al. "KiDS-1000 Cosmology: Multi-probe weak gravitational lensing and spectroscopic galaxy clustering constraints". In: *A&A* 646, A140 (Feb. 2021), A140. DOI: 10.1051/0004-6361/202039063. arXiv: 2007.15632 [astro-ph.CO].

[131] Chiaki Hikage et al. "Cosmology from cosmic shear power spectra with Subaru Hyper Suprime-Cam first-year data". In: *PASJ* 71.2, 43 (Apr. 2019), p. 43. DOI: 10.1093/pasj/psz010. arXiv: 1809.09148 [astro-ph.CO].

[132] H. Hildebrandt, M. Viola, and C. Heymans et al. "KiDS-450: cosmological parameter constraints from tomographic weak gravitational lensing". In: *MNRAS* 465.2 (Feb. 2017), pp. 1454–1498. DOI: 10.1093/mnras/stw2805. arXiv: 1606.05338 [astro-ph.CO].

[133] Eric Hivon et al. "MASTER of the Cosmic Microwave Background Anisotropy Power Spectrum: A Fast Method for Statistical Analysis of Large and Complex Cosmic Microwave Background Data Sets". In: *The Astrophysical Journal* 567.1 (Mar. 2002), pp. 2–17. DOI: 10.1086/338126. arXiv: astro-ph/0105302 [astro-ph].

[134] H. Hoekstra and B. Jain. "Weak Gravitational Lensing and Its Cosmological Applications". In: *Annual Review of Nuclear and Particle Science* 58 (Nov. 2008), pp. 99–123. DOI: 10.1146/annurev.nucl.58.110707.171151. arXiv: 0805.0139.

[135] Philip F. Hopkins. "A new class of accurate, mesh-free hydrodynamic simulation methods". In: *Monthly Notices of the Royal Astronomical Society* 450.1 (Apr. 2015), pp. 53–110. ISSN: 1365-2966. DOI: 10.1093/mnras/stv195. URL: http://dx.doi.org/10.1093/mnras/stv195.

[136] Benjamin Horowitz and Peter Melchior. "Plausible Adversarial Attacks on Direct Parameter Inference Models in Astrophysics". In: *arXiv preprint arXiv:2211.14788* (2022).

[137] Wayne Hu. "Dark energy and matter evolution from lensing tomography". In: *Phys. Rev. D* 66.8, 083515 (Oct. 2002), p. 083515. DOI: 10.1103/PhysRevD.66.083515. arXiv: astro-ph/0208093 [astro-ph].

[138] Hung-Jin Huang et al. "Modelling baryonic physics in future weak lensing surveys". In: *Monthly Notices of the Royal Astronomical Society* 488.2 (2019), pp. 1652–1678.

[139] Dragan Huterer. "Weak lensing and dark energy". In: *Phys. Rev. D* 65.6, 063001 (Mar. 2002), p. 063001. DOI: 10.1103/PhysRevD.65.063001. arXiv: astro-ph/0106399 [astro-ph].

[140] Sergey Ioffe and Christian Szegedy. "Batch normalization: Accelerating deep network training by reducing internal covariate shift". In: *International conference on machine learning*. PMLR. 2015, pp. 448–456.

[141] Željko Ivezić et al. "LSST: from science drivers to reference design and anticipated data products". In: *The Astrophysical Journal* 873.2 (2019), p. 111.

[142] A. Izard, M. Crocce, and P. Fosalba. "ICE-COLA: towards fast and accurate synthetic galaxy catalogues optimizing a quasi-N-body method". In: *MNRAS* 459 (July 2016), pp. 2327–2341. DOI: 10.1093/mnras/stw797. arXiv: 1509.04685.

[143] Bhuvnesh Jain and Uroš Seljak. "Cosmological Model Predictions for Weak Lensing: Linear and Nonlinear Regimes". In: *ApJ* 484.2 (July 1997), pp. 560–573. DOI: 10.1086/304372. arXiv: astro-ph/9611077 [astro-ph].

[144] Bhuvnesh Jain and Ludovic Van Waerbeke. "Statistics of Dark Matter Halos from Gravitational Lensing". In: *The Astrophysical Journal Letters* 530.1 (Feb. 2000), pp. L1–L4. DOI: 10.1086/312480. arXiv: astro-ph/9910459 [astro-ph].

[145] M. Jarvis, G. Bernstein, and B. Jain. "The skewness of the aperture mass statistic". In: *MNRAS* 352.1 (July 2004), pp. 338–352. DOI: 10.1111/j.1365-2966.2004.07926.x. arXiv: astro-ph/0307393 [astro-ph].

[146] Jens Jasche and Francisco S. Kitaura. "Fast Hamiltonian sampling for large-scale structure inference". In: *MNRAS* 407.1 (Sept. 2010), pp. 29–42. DOI: 10.1111/j.1365-2966.2010.16897.x. arXiv: 0911.2496 [astro-ph.CO].

[147] Jens Jasche and Benjamin D. Wandelt. "Bayesian physical reconstruction of initial conditions from large-scale structure surveys". In: *Monthly Notices of the Royal Astronomical Society* 432.2 (June 2013), pp. 894–913. DOI: 10.1093/mnras/stt449. arXiv: 1203.3639 [astro-ph.CO].

[148] N. Jeffrey et al. "Dark Energy Survey Year 3 results: likelihood-free, simulation-based $w$CDM inference with neural compression of weak-lensing map statistics". In: *arXiv e-prints*, arXiv:2403.02314 (Mar. 2024), arXiv:2403.02314. DOI: 10.48550/arXiv.2403.02314. arXiv: 2403.02314 [astro-ph.CO].

[149] Niall Jeffrey, Justin Alsing, and François Lanusse. "Likelihood-free inference with neural compression of DES SV weak lensing map statistics". In: *Monthly Notices of the Royal Astronomical Society* 501.1 (Feb. 2021), pp. 954–969. DOI: 10.1093/mnras/staa3594. arXiv: 2009.08459 [astro-ph.CO].

[150] A. Jenkins et al. "Evolution of Structure in Cold Dark Matter Universes". In: *The Astrophysical Journal* 499.1 (May 1998), pp. 20–40. ISSN: 1538-4357. DOI: 10.1086/305615. URL: http://dx.doi.org/10.1086/305615.

[151] Adrian Jenkins. "Second-order Lagrangian perturbation theory initial conditions for resimulations". In: *MNRAS* 403.4 (Apr. 2010), pp. 1859–1872. DOI: 10.1111/j.1365-2966.2010.16259.x. arXiv: 0910.0258 [astro-ph.CO].

[152] S. Joudaki et al. "KiDS-450 + 2dFLenS: Cosmological parameter constraints from weak gravitational lensing tomography and overlapping redshift-space galaxy clustering". In: *MNRAS* 474 (Mar. 2018), pp. 4894–4924. DOI: 10.1093/mnras/stx2820. arXiv: 1707.06627.

[153] Shahab Joudaki et al. "CFHTLenS revisited: assessing concordance with Planck including astrophysical systematics". In: *Monthly Notices of the Royal Astronomical Society* (2016), stw2665.

[154] H M Dipu Kabir et al. *SpinalNet: Deep Neural Network with Gradual Input*. 2022. arXiv: 2007.03347 [cs.CV].

[155] T. Kacprzak et al. "Cosmology constraints from shear peak statistics in Dark Energy Survey Science Verification data". In: *Monthly Notices of the Royal Astronomical Society* 463.4 (Dec. 2016), pp. 3653–3673. DOI: 10.1093/mnras/stw2070. arXiv: 1603.05040 [astro-ph.CO].

[156] Mahdi Karami et al. "Invertible Convolutional Flow". In: *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*. Ed. by Hanna M. Wallach et al. 2019, pp. 5636–5646. URL: https://proceedings.neurips.cc/paper/2019/hash/b1f62fa99de9f27a048344d55c5ef7a6-Abstract.html.

[157] Tero Karras et al. "Analyzing and improving the image quality of stylegan". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020, pp. 8110–8119.

[158] Maurice George Kendall and Alan Stuart. *The Advanced Theory of Statistics*. Charles Griffin and Co., Ltd., London, 1946.

[159] M. Kilbinger. "Cosmology with cosmic shear observations: a review". In: *Reports on Progress in Physics* 78.8, 086901 (July 2015), p. 086901. DOI: 10.1088/0034-4885/78/8/086901. arXiv: 1411.0115.

[160] Diederik Kingma et al. "Variational Diffusion Models". In: *Advances in Neural Information Processing Systems*. Ed. by M. Ranzato et al. Vol. 34. Curran Associates, Inc., 2021, pp. 21696–21707. URL: https://proceedings.neurips.cc/paper_files/paper/2021/file/b578f2a52a0229873fefc2a4b06377fa-Paper.pdf.

[161] Diederik P. Kingma and Prafulla Dhariwal. "Glow: Generative Flow with Invertible 1x1 Convolutions". In: *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montréal, Canada*. Ed. by Samy Bengio et al. 2018, pp. 10236–10245.

[162] Diederik P. Kingma and Max Welling. "Auto-Encoding Variational Bayes". In: *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings*. Ed. by Yoshua Bengio and Yann LeCun. 2014. URL: http://arxiv.org/abs/1312.6114.

[163] F. -S. Kitaura. "The initial conditions of the universe from constrained simulations." In: *Monthly Notices of the Royal Astronomical Society* 429 (Feb. 2013), pp. L84–L88. DOI: 10.1093/mnrasl/sls029. arXiv: 1203.4184 [astro-ph.CO].

[164] F.-S. Kitaura and S. Heß. "Cosmological structure formation with augmented Lagrangian perturbation theory". In: *MNRAS* 435 (Aug. 2013), pp. L78–L82. DOI: 10.1093/mnrasl/slt101. arXiv: 1212.3514.

[165] F.-S. Kitaura et al. "The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: mock galaxy catalogues for the BOSS Final Data Release". In: *MNRAS* 456 (Mar. 2016), pp. 4156–4173. DOI: 10.1093/mnras/stv2826. arXiv: 1509.06400.

[166] T. D. Kitching, A. F. Heavens, and J. Alsing et al. "3D cosmic shear: cosmology from CFHTLenS". In: *MNRAS* 442.2 (Aug. 2014), pp. 1326–1349. DOI: 10.1093/mnras/stu934. arXiv: 1401.6842 [astro-ph.CO].

[167] T. D. Kitching, A. F. Heavens, and L. Miller. "3D photometric cosmic shear". In: *MNRAS* 413.4 (June 2011), pp. 2923–2934. DOI: 10.1111/j.1365-2966.2011.18369.x. arXiv: 1007.2953 [astro-ph.CO].

[168] A. Klypin and F. Prada. "Dark matter statistics for large galaxy catalogues: power spectra and covariance matrices". In: *MNRAS* 478 (Aug. 2018), pp. 4602–4621. DOI: 10.1093/mnras/sty1340. arXiv: 1701.05690.

[169] Anatoly Klypin et al. "Galaxies in N-Body Simulations: Overcoming the Overmerging Problem". In: *ApJ* 516.2 (May 1999), pp. 530–551. DOI: 10.1086/307122. arXiv: astro-ph/9708191 [astro-ph].

[170] M Knabenhans et al. "Euclid preparation: IX. EuclidEmulator2 – power spectrum emulation with massive neutrinos and self-consistent dark energy perturbations". In: *Monthly Notices of the Royal Astronomical Society* 505.2 (May 2021), pp. 2840–2869. ISSN: 1365-2966. DOI: 10.1093/mnras/stab1366. URL: http://dx.doi.org/10.1093/mnras/stab1366.

[171] Doogesh Kodi Ramanah et al. "Super-resolution emulator of cosmological simulations using deep physical models". In: *Monthly Notices of the Royal Astronomical Society* 495.4 (2020), pp. 4227–4236.

[172] Frederic Koehler, Viraj Mehta, and Andrej Risteski. "Representational aspects of depth and conditioning in normalizing flows". In: *International Conference on Machine Learning*. PMLR. 2021, pp. 5628–5636.

[173] Fabian Köhlinger et al. "A direct measurement of tomographic lensing power spectra from CFHTLenS". In: *Monthly Notices of the Royal Astronomical Society* 456.2 (2016), pp. 1508–1527.

[174] Fabian Köhlinger et al. "KiDS-450: The tomographic weak lensing power spectrum and constraints on cosmological parameters". In: *Monthly Notices of the Royal Astronomical Society* 471.4 (2017), pp. 4412–4435.

[175] Jan M Kratochvil et al. "Probing cosmology with weak lensing Minkowski functionals". In: *Physical Review D* 85.10 (2012), p. 103513.

[176] Jan M. Kratochvil, Zoltán Haiman, and Morgan May. "Probing cosmology with weak lensing peak counts". In: *Physical Review D* 81.4, 043519 (Feb. 2010), p. 043519. DOI: 10.1103/PhysRevD.81.043519. arXiv: 0907.0486 [astro-ph.CO].

[177] E. Krause et al. "Dark Energy Survey Year 1 Results: Multi-Probe Methodology and Simulated Likelihood Analyses". In: *arXiv e-prints*, arXiv: 1706.09359 (June 2017), arXiv: 1706.09359. arXiv: 1706.09359 [astro-ph.CO].

[178] E. Krause et al. "Dark Energy Survey Year 3 Results: Multi-Probe Modeling Strategy and Validation". In: *arXiv e-prints*, arXiv:2105.13548 (May 2021), arXiv:2105.13548. arXiv: 2105.13548 [astro-ph.CO].

[179] Rene Laureijs et al. "Euclid definition study report". In: *arXiv preprint arXiv:1110.3193* (2011).

[180] F. Leclercq et al. "One-point remapping of Lagrangian perturbation theory in the mildly non-linear regime of cosmic structure formation". In: *J. Cosmology Astropart. Phys.* 11, 048 (Nov. 2013), p. 048. DOI: 10.1088/1475-7516/2013/11/048. arXiv: 1305.4642.

[181] M. Levi et al. "The DESI Experiment, a whitepaper for Snowmass 2013". In: *arXiv e-prints* (Aug. 2013). arXiv: 1308.0847 [astro-ph.CO].

[182] Antony Lewis and Anthony Challinor. *CAMB: Code for Anisotropies in the Microwave Background*. Astrophysics Source Code Library, record ascl:1102.026. Feb. 2011. ascl: 1102.026.

[183] Xiangchong Li et al. "Hyper Suprime-Cam Year 3 results: Cosmology from cosmic shear two-point correlation functions". In: *Phys. Rev. D* 108.12, 123518 (Dec. 2023), p. 123518. DOI: 10.1103/PhysRevD.108.123518. arXiv: 2304.00702 [astro-ph.CO].

[184] Yin Li et al. "AI-assisted super-resolution cosmological simulations". In: *arXiv preprint arXiv:2010.06608* 2020 (2020).

[185] Yingzhen Li, John Bradshaw, and Yash Sharma. "Are Generative Classifiers More Robust to Adversarial Attacks?" In: *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*. Ed. by Kamalika Chaudhuri and Ruslan Salakhutdinov. Vol. 97. Proceedings of Machine Learning Research. PMLR, 2019, pp. 3804–3814. URL: http://proceedings.mlr.press/v97/li19a.html.

[186] Zack Li, Jia Liu, and José Manuel Zorrilla Matilla et al. "Constraining neutrino mass with tomographic weak lensing peak counts". In: *Phys. Rev. D* 99.6, 063527 (Mar. 2019), p. 063527. DOI: 10.1103/PhysRevD.99.063527. arXiv: 1810.01781 [astro-ph.CO].

[187] Chieh-An Lin and Martin Kilbinger. "A new model to predict weak-lensing peak counts. II. Parameter constraint strategies". In: *A&A* 583, A70 (Nov. 2015), A70. DOI: 10.1051/0004-6361/201526659. arXiv: 1506.01076 [astro-ph.CO].

[188] Florian List, Ishaan Bhat, and Geraint F Lewis. "A black box for dark sector physics: Predicting dark matter annihilation feedback with conditional GANs". In: *Monthly Notices of the Royal Astronomical Society* 490.3 (2019), pp. 3134–3143.

[189] Florian List and Oliver Hahn. "Perturbation-theory informed integrators for cosmological simulations". In: *Journal of Computational Physics* 513, 113201 (Sept. 2024), p. 113201. DOI: 10.1016/j.jcp.2024.113201. arXiv: 2301.09655 [astro-ph.CO].

[190] Hao Liu and Pieter Abbeel. "Hybrid Discriminative-Generative Training via Contrastive Learning". In: *CoRR* abs/2007.09070 (2020). arXiv: 2007.09070. URL: https://arxiv.org/abs/2007.09070.

[191] Jia Liu, Simeon Bird, and José Manuel Zorrilla Matilla et al. "MassiveNuS: cosmological massive neutrino simulations". In: *J. Cosmology Astropart. Phys.* 2018.3, 049 (Mar. 2018), p. 049. DOI: 10.1088/1475-7516/2018/03/049. arXiv: 1711.10524 [astro-ph.CO].

[192] Jia Liu, Andrea Petri, and Zoltán Haiman et al. "Cosmology constraints from the weak lensing peak counts and the power spectrum in CFHTLenS data". In: *Phys. Rev. D* 91.6, 063507 (Mar. 2015), p. 063507. DOI: 10.1103/PhysRevD.91.063507. arXiv: 1412.0757 [astro-ph.CO].

[193] Xiangkun Liu, Chuzhong Pan, and Ran Li et al. "Cosmological constraints from weak lensing peak statistics with Canada-France-Hawaii Telescope Stripe 82 Survey". In: *MNRAS* 450.3 (July 2015), pp. 2888–2902. DOI: 10.1093/mnras/stv784. arXiv: 1412.3683 [astro-ph.CO].

[194] Xiangkun Liu et al. "Cosmological studies from HSC-SSP tomographic weak-lensing peak abundances". In: *Monthly Notices of the Royal Astronomical Society* 519.1 (2023), pp. 594–612.

[195] LSST Science Collaboration et al. "LSST Science Book, Version 2.0". In: *arXiv e-prints*, arXiv:0912.0201 (Dec. 2009), arXiv:0912.0201. arXiv: 0912.0201 [astro-ph.IM].

[196] Tianhuan Lu and Zoltán Haiman. "The impact of baryons on cosmological inference from weak lensing statistics". In: *Monthly Notices of the Royal Astronomical Society* 506.3 (Sept. 2021), pp. 3406–3417. DOI: 10.1093/mnras/stab1978. arXiv: 2104.04165 [astro-ph.CO].

[197] Tianhuan Lu, Zoltán Haiman, and Xiangchong Li. "Cosmological constraints from HSC survey first-year data using deep learning". In: *arXiv preprint arXiv:2301.01354* (2023).

[198] Tianhuan Lu, Zoltán Haiman, and José Manuel Zorrilla Matilla. "Simultaneously constraining cosmology and baryonic physics via deep learning from weak lensing". In: *Monthly Notices of the Royal Astronomical Society* 511.1 (Mar. 2022), pp. 1518–1528. DOI: 10.1093/mnras/stac161. arXiv: 2109.11060 [astro-ph.CO].

[199] N. MacCrann et al. "DES Y1 Results: validating cosmological parameter estimation using simulated Dark Energy Surveys". In: *MNRAS* 480 (Nov. 2018), pp. 4614–4635. DOI: 10.1093/mnras/sty1899. arXiv: 1803.09795.

[200] M. Maciejewski et al. "Phase-space structures - II. Hierarchical Structure Finder". In: *MNRAS* 396.3 (July 2009), pp. 1329–1348. DOI: 10.1111/j.1365-2966.2009.14825.x. arXiv: 0812.0288 [astro-ph].

[201] Durjoy Sen Maitra, Ujjwal Bhattacharya, and Swapan K. Parui. "CNN based common approach to handwritten character recognition of multiple scripts". In: *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*. 2015, pp. 1021–1025. DOI: 10.1109/ICDAR.2015.7333916.

[202] T. Lucas Makinen et al. "Lossless, scalable implicit likelihood inference for cosmological fields". In: *Journal of Cosmology and Astroparticle Physics* 2021.11, 049 (Nov. 2021), p. 049. DOI: 10.1088/1475-7516/2021/11/049. arXiv: 2107.07405 [astro-ph.CO].

[203] Stéphane Mallat. "Group invariant scattering". In: *Communications on Pure and Applied Mathematics* 65.10 (2012), pp. 1331–1398.

[204] Stephane G Mallat. "A theory for multiresolution signal decomposition: the wavelet representation". In: *IEEE transactions on pattern analysis and machine intelligence* 11.7 (1989), pp. 674–693.

[205] Marc Manera et al. "The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: a large sample of mock galaxy catalogues". In: *MNRAS* 428.2 (Jan. 2013), pp. 1036–1054. DOI: 10.1093/mnras/sts084. arXiv: 1203.6609 [astro-ph.CO].

[206] Laura Marian, Robert E. Smith, and Gary M. Bernstein. "The cosmology dependence of weak lensing cluster counts". In: *The Astrophysical Journal* 698.1 (May 2009), pp. L33–L36. ISSN: 1538-4357. DOI: 10.1088/0004-637x/698/1/l33. URL: http://dx.doi.org/10.1088/0004-637X/698/1/L33.

[207] Laura Marian, Robert E. Smith, and Stefan Hilbert et al. "Optimized detection of shear peaks in weak lensing maps". In: *MNRAS* 423.2 (June 2012), pp. 1711–1725. DOI: 10.1111/j.1365-2966.2012.20992.x. arXiv: 1110.4635 [astro-ph.CO].

[208] Federico Marinacci et al. "First results from the IllustrisTNG simulations: radio haloes and magnetic fields". In: *Monthly Notices of the Royal Astronomical Society* 480.4 (2018), pp. 5113–5139.

[209] Nicolas Martinet et al. "KiDS-450: cosmological constraints from weak-lensing peak statistics–II: Inference from shear peaks using N-body simulations". In: *Monthly Notices of the Royal Astronomical Society* 474.1 (2018), pp. 712–730.

[210] José Manuel Zorrilla Matilla et al. "Interpreting deep learning models for weak lensing". In: *Phys. Rev. D* 102 (12 Dec. 2020), p. 123506. DOI: 10.1103/PhysRevD.102.123506. URL: https://link.aps.org/doi/10.1103/PhysRevD.102.123506.

[211] M. Maturi, C. Fedeli, and L. Moscardini. "Imprints of primordial non-Gaussianity on the number counts of cosmic shear peaks". In: *MNRAS* 416.4 (Oct. 2011), pp. 2527–2538. DOI: 10.1111/j.1365-2966.2011.18958.x. arXiv: 1101.4175 [astro-ph.CO].

[212] S. McAlpine et al. "The EAGLE simulations of galaxy formation: Public release of halo and galaxy catalogues". In: *Astronomy and Computing* 15 (Apr. 2016), pp. 72–89. DOI: 10.1016/j.ascom.2016.02.004. arXiv: 1510.01320 [astro-ph.GA].

[213] Andrew McCallum et al. "Multi-conditional learning: Generative/discriminative training for clustering and classification". In: *AAAI*. Vol. 1. 2006, p. 6.

[214] I. G. McCarthy et al. "Gas expulsion by quasar-driven winds as a solution to the overcooling problem in galaxy groups and clusters". In: *MNRAS* 412 (Apr. 2011), pp. 1965–1984. DOI: 10.1111/j.1365-2966.2010.18033.x. arXiv: 1008.4799 [astro-ph.CO].

[215] Ian G. McCarthy et al. "The BAHAMAS project: calibrated hydrodynamical simulations for large-scale structure cosmology". In: *MNRAS* 465.3 (Mar. 2017), pp. 2936–2965. DOI: 10.1093/mnras/stw2792. arXiv: 1603.02702 [astro-ph.CO].

[216] Ian G. McCarthy et al. "The BAHAMAS project: the CMB-large-scale structure tension and the roles of massive neutrinos and galaxy formation". In: *MNRAS* 476.3 (May 2018), pp. 2999–3030. DOI: 10.1093/mnras/sty377. arXiv: 1712.02411 [astro-ph.CO].

[217] Thomas McClintock et al. "The Aemulus Project. II. Emulating the Halo Mass Function". In: *ApJ* 872.1, 53 (Feb. 2019), p. 53. DOI: 10.3847/1538-4357/aaf568. arXiv: 1804.05866 [astro-ph.CO].

[218] Jason D. McEwen, Christopher G. R. Wallis, and Augustine N. Mavor-Parker. "Scattering Networks on the Sphere for Scalable and Rotationally Equivariant Spherical CNNs". In: *arXiv e-prints*, arXiv:2102.02828 (Feb. 2021), arXiv:2102.02828. arXiv: 2102.02828 [cs.CV].

[219] A. J. Mead et al. "HMCODE-2020: improved modelling of non-linear cosmological power spectra with baryonic feedback". In: *MNRAS* 502.1 (Mar. 2021), pp. 1401–1422. DOI: 10.1093/mnras/stab082. arXiv: 2009.01858 [astro-ph.CO].

[220] K. R. Mecke, T. Buchert, and H. Wagner. "Robust morphological measures for large-scale structure in the Universe". In: *Astronomy and Astrophysics* 288 (Aug. 1994), pp. 697–704. DOI: 10.48550/arXiv.astro-ph/9312028. arXiv: astro-ph/9312028 [astro-ph].

[221] Michaël Michaux et al. "Accurate initial conditions for cosmological N-body simulations: minimizing truncation and discreteness errors". In: *MNRAS* 500.1 (Jan. 2021), pp. 663–683. DOI: 10.1093/mnras/staa3149. arXiv: 2008.09588 [astro-ph.CO].

[222] C. Modi, F. Lanusse, and U. Seljak. "FlowPM: Distributed TensorFlow implementation of the FastPM cosmological N-body solver". In: *Astronomy and Computing* 37, 100505 (Oct. 2021), p. 100505. DOI: 10.1016/j.ascom.2021.100505. arXiv: 2010.11847 [astro-ph.CO].

[223] C. Modi et al. "Intensity mapping with neutral hydrogen and the Hidden Valley simulations". In: *arXiv e-prints* (Apr. 2019). arXiv: 1904.11923.

[224] Chirag Modi, Yu Feng, and Uroš Seljak. "Cosmological reconstruction from galaxy light: neural network based light-matter connection". In: *Journal of Cosmology and Astroparticle Physics* 2018.10, 028 (Oct. 2018), p. 028. DOI: 10.1088/1475-7516/2018/10/028. arXiv: 1805.02247 [astro-ph.CO].

[225] Chirag Modi et al. "Sensitivity Analysis of Simulation-Based Inference for Galaxy Clustering". In: *arXiv e-prints*, arXiv:2309.15071 (Sept. 2023), arXiv:2309.15071. DOI: 10.48550/arXiv.2309.15071. arXiv: 2309.15071 [astro-ph.CO].

[226] Nayantara Mudur and Douglas P Finkbeiner. "Can denoising diffusion probabilistic models generate realistic astrophysical fields?" In: *arXiv preprint arXiv:2211.12444* (2022).

[227] Mustafa Mustafa, Deborah Bard, and Wahid Bhimji et al. "CosmoGAN: creating high-fidelity weak lensing convergence maps using Generative Adversarial Networks". In: *Computational Astrophysics and Cosmology* 6.1, 1 (May 2019), p. 1. DOI: 10.1186/s40668-019-0029-9. arXiv: 1706.02390 [astro-ph.IM].

[228] Jill P Naiman et al. "First results from the IllustrisTNG simulations: A tale of two elements–chemical evolution of magnesium and europium". In: *Monthly Notices of the Royal Astronomical Society* 477.1 (2018), pp. 1206–1224.

[229] J. F. Navarro, C. S. Frenk, and S. D. M. White. "A Universal Density Profile from Hierarchical Clustering". In: *ApJ* 490 (Dec. 1997), pp. 493–508. DOI: 10.1086/304888. eprint: astro-ph/9611107.

[230] J. F. Navarro, C. S. Frenk, and S. D. M. White. "The Structure of Cold Dark Matter Halos". In: *ApJ* 462 (May 1996), p. 563. DOI: 10.1086/177173. eprint: astro-ph/9508025.

[231] D. Nelson et al. "The illustris simulation: Public data release". In: *Astronomy and Computing* 13 (Nov. 2015), pp. 12–37. DOI: 10.1016/j.ascom.2015.09.003. arXiv: 1504.00362.

[232] Dylan Nelson et al. "First results from the IllustrisTNG simulations: the galaxy colour bimodality". In: *Monthly Notices of the Royal Astronomical Society* 475.1 (2018), pp. 624–647.

[233] M. C. Neyrinck. "Truthing the stretch: non-perturbative cosmological realizations with multiscale spherical collapse". In: *MNRAS* 455 (Jan. 2016), pp. L11–L15. DOI: 10.1093/mnrasl/slv141. arXiv: 1503.07534.

[234] Mark C. Neyrinck, István Szapudi, and Alexander S. Szalay. "Rejuvenating the Matter Power Spectrum: Restoring Information with a Logarithmic Density Mapping". In: *The Astrophysical Journal Letters* 698.2 (June 2009), pp. L90–L93. DOI: 10.1088/0004-637X/698/2/L90. arXiv: 0903.4693 [astro-ph.CO].

[235] Andrew Y Ng and Michael I Jordan. "On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes". In: *Advances in neural information processing systems*. 2002, pp. 841–848.

[236] Yueying Ni et al. "The ASTRID simulation: the evolution of supermassive black holes". In: *MNRAS* 513.1 (June 2022), pp. 670–692. DOI: 10.1093/mnras/stac351. arXiv: 2110.14154 [astro-ph.GA].

[237] Yueying Ni et al. "The CAMELS project: Expanding the galaxy formation model space with new ASTRID and 28-parameter TNG and SIMBA suites". In: *arXiv e-prints*, arXiv: 2304.02096 (Apr. 2023), arXiv: 2304.02096. DOI: 10.48550/arXiv.2304.02096. arXiv: 2304.02096 [astro-ph.CO].

[238] Aaron van den Oord et al. "Conditional image generation with pixelcnn decoders". In: *arXiv preprint arXiv:1606.05328* (2016).

[239] George Papamakarios, Iain Murray, and Theo Pavlakou. "Masked Autoregressive Flow for Density Estimation". In: *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*. Ed. by Isabelle Guyon et al. 2017, pp. 2338–2347.

[240] Core Francisco Park et al. "Quantification of high dimensional non-Gaussianities and its implication to Fisher analysis in cosmology". In: *arXiv preprint arXiv:2204.05435* (2022).

[241] Carolina Parroni et al. "Going deep with Minkowski functionals of convergence maps". In: *Astronomy and Astrophysics* 633, A71 (Jan. 2020), A71. DOI: 10.1051/0004-6361/201935988. arXiv: 1911.06243 [astro-ph.CO].

[242] John A. Peacock. *Cosmological Physics*. 1999.

[243] Christian Pedersen et al. "An emulator for the Lyman-$\alpha$ forest in beyond-$\Lambda$CDM cosmologies". In: *Journal of Cosmology and Astroparticle Physics* 2021.05 (May 2021), p. 033. ISSN: 1475-7516. DOI: 10.1088/1475-7516/2021/05/033. URL: http://dx.doi.org/10.1088/1475-7516/2021/05/033.

[244] P. J. E. Peebles. *Principles of Physical Cosmology*. Princeton University Press, 1993.

[245] P. J. E. Peebles. *The large-scale structure of the universe*. 1980.

[246] P. J. E. Peebles and E. J. Groth. "Statistical analysis of catalogs of extragalactic objects. V. Three-point correlation function for the galaxy distribution in the Zwicky catalog." In: *The Astrophysical Journal* 196 (Feb. 1975), pp. 1–11. DOI: 10.1086/153390.

[247] Austin Peel, Chieh-An Lin, and François Lanusse et al. "Cosmological constraints with weak-lensing peak counts and second-order statistics in a large-field survey". In: *A&A* 599, A79 (Mar. 2017), A79. DOI: 10.1051/0004-6361/201629928. arXiv: 1612.02264 [astro-ph.CO].

[248] Ue-Li Pen, Tongjie Zhang, and Ludovic van Waerbeke et al. "Detection of Dark Matter Skewness in the VIRMOS-DESCART Survey: Implications for $\Omega_0$". In: *ApJ* 592.2 (Aug. 2003), pp. 664–673. DOI: 10.1086/375734. arXiv: astro-ph/0302031 [astro-ph].

[249] Nathanaël Perraudin, Sandro Marcon, and Aurelien Lucchi et al. "Emulation of cosmological mass maps with conditional generative adversarial networks". In: *arXiv e-prints*, arXiv:2004.08139 (Apr. 2020), arXiv:2004.08139. arXiv: 2004.08139 [astro-ph.CO].

[250] A. Petri. "Mocking the weak lensing universe: The LensTools Python computing package". In: *Astronomy and Computing* 17 (Oct. 2016), pp. 73–79. DOI: 10.1016/j.ascom.2016.06.001. arXiv: 1606.01903 [astro-ph.CO].

[251] Andrea Petri, Zoltán Haiman, and Lam Hui et al. "Cosmology with Minkowski functionals and moments of the weak lensing convergence field". In: *Phys. Rev. D* 88.12, 123002 (Dec. 2013), p. 123002. DOI: 10.1103/PhysRevD.88.123002. arXiv: 1309.4460 [astro-ph.CO].

[252] Annalisa Pillepich et al. "First results from the IllustrisTNG simulations: the stellar mass content of groups and clusters of galaxies". In: *Monthly Notices of the Royal Astronomical Society* 475.1 (2018), pp. 648–675.

[253] Annalisa Pillepich et al. "Simulating galaxy formation with the IllustrisTNG model". In: *Monthly Notices of the Royal Astronomical Society* 473.3 (2018), pp. 4077–4106.

[254] Sandrine Pires, Adrienne Leonard, and Jean-Luc Starck. "Cosmological constraints from the capture of non-Gaussianity in weak lensing data". In: *MNRAS* 423.1 (June 2012), pp. 983–992. DOI: 10.1111/j.1365-2966.2012.20940.x. arXiv: 1203.2877 [astro-ph.CO].

[255] Alice Pisani et al. "Cosmic voids: a novel probe to shed light on our Universe". In: *Bulletin of the American Astronomical Society* 51.3, 40 (May 2019), p. 40. DOI: 10.48550/arXiv.1903.05161. arXiv: 1903.05161 [astro-ph.CO].

[256] Planck Collaboration et al. "Planck 2015 results. XIII. Cosmological parameters". In: *Astronomy and Astrophysics* 594, A13 (Sept. 2016), A13. DOI: 10.1051/0004-6361/201525830. arXiv: 1502.01589 [astro-ph.CO].

[257] Planck Collaboration et al. "Planck 2018 results. VI. Cosmological parameters". In: *A&A* 641, A6 (Sept. 2020), A6. DOI: 10.1051/0004-6361/201833910. arXiv: 1807.06209 [astro-ph.CO].

[258] Natalia Porqueres, Alan Heavens, and Daniel Mortlock et al. "Bayesian forward modelling of cosmic shear data". In: *arXiv e-prints*, arXiv:2011.07722 (Nov. 2020), arXiv:2011.07722. arXiv: 2011.07722 [astro-ph.CO].

[259] Natalia Porqueres et al. "Lifting weak lensing degeneracies with a field-based likelihood". In: *MNRAS* 509.3 (Jan. 2022), pp. 3194–3202. DOI: 10.1093/mnras/stab3234. arXiv: 2108.04825 [astro-ph.CO].

[260] K. Price, Rainer Storn, and J. Lampinen. *Differential Evolution-A Practical Approach to Global Optimization*. Vol. 141. Jan. 2005. DOI: 10.1007/3-540-31306-0.

[261] Thomas Quinn et al. "Time stepping N-body simulations". In: *arXiv e-prints*, astro-ph/9710043 (Oct. 1997), astro–ph/9710043. DOI: 10.48550/arXiv.astro-ph/9710043. arXiv: astro-ph/9710043 [astro-ph].

[262] Alec Radford, Luke Metz, and Soumith Chintala. "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks". In: *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings*. Ed. by Yoshua Bengio and Yann LeCun. 2016. URL: http://arxiv.org/abs/1511.06434.

[263] Rajat Raina et al. "Classification with hybrid generative/discriminative models". In: *Advances in neural information processing systems* 16 (2003).

[264] Doogesh Kodi Ramanah, Tom Charnock, and Guilhem Lavaux. "Painting halos from cosmic density fields of dark matter with physically motivated neural networks". In: *Physical Review D* 100.4 (2019), p. 043515.

[265] Yann Rasera et al. "Cosmic-variance limited Baryon Acoustic Oscillations from the DEUS-FUR ΛCDM simulation". In: *Monthly Notices of the Royal Astronomical Society* 440.2 (2014), pp. 1420–1434.

[266] Carl Rasmussen et al. "Gaussian Processes in Machine Learning". In: *Advanced Lectures on Machine Learning: ML Summer Schools 2003, Canberra, Australia, February 2 - 14, 2003, Tübingen, Germany, August 4 - 16, 2003, Revised Lectures, 63-71 (2004)* 3176 (Sept. 2004). DOI: 10.1007/978-3-540-28650-9_4.

[267] Ali Razavi, Aaron van den Oord, and Oriol Vinyals. "Generating diverse high-fidelity images with vq-vae-2". In: *Advances in neural information processing systems*. 2019, pp. 14866–14876.

[268] Danilo Jimenez Rezende and Shakir Mohamed. "Variational Inference with Normalizing Flows". In: *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015*. Ed. by Francis R. Bach and David M. Blei. Vol. 37. JMLR Workshop and Conference Proceedings. JMLR.org, 2015, pp. 1530–1538. URL: http://proceedings.mlr.press/v37/rezende15.html.

[269] Danilo Jimenez Rezende, Shakir Mohamed, and Daan Wierstra. "Stochastic Backpropagation and Approximate Inference in Deep Generative Models". In: *Proceedings of the 31th International Conference on Machine Learning, ICML 2014, Beijing, China, 21-26 June 2014*. Vol. 32. JMLR Workshop and Conference Proceedings. JMLR.org, 2014, pp. 1278–1286. URL: http://proceedings.mlr.press/v32/rezende14.html.

[270] Dezső Ribli et al. "Weak lensing cosmology with convolutional neural networks on noisy data". In: *Monthly Notices of the Royal Astronomical Society* 490.2 (Dec. 2019), pp. 1843–1860. DOI: 10.1093/mnras/stz2610. arXiv: 1902.03663 [astro-ph.CO].

[271] Adam G. Riess et al. "Large Magellanic Cloud Cepheid Standards Provide a 1% Foundation for the Determination of the Hubble Constant and Stronger Evidence for Physics beyond ΛCDM". In: *ApJ* 876.1, 85 (May 2019), p. 85. DOI: 10.3847/1538-4357/ab1422. arXiv: 1903.07603 [astro-ph.CO].

[272] Keir K. Rogers and Hiranya V. Peiris. "General framework for cosmological dark matter bounds using ¡mml:math xmlns:mml="http://www.w3.org/1998/Math/MathML" display="inline"¿¡mml:mi¿N¡/mml:mi¿¡/mml:math¿ -body simulations". In: *Physical Review D* 103.4 (Feb. 2021). ISSN: 2470-0029. DOI: 10.1103/physrevd.103.043526. URL: http://dx.doi.org/10.1103/PhysRevD.103.043526.

[273] Keir K. Rogers et al. "Bayesian emulator optimisation for cosmology: application to the Lyman-alpha forest". In: *Journal of Cosmology and Astroparticle Physics* 2019.02 (Feb. 2019), pp. 031–031. ISSN: 1475-7516. DOI: 10.1088/1475-7516/2019/02/031. URL: http://dx.doi.org/10.1088/1475-7516/2019/02/031.

[274] Adam Rouhiainen, Utkarsh Giri, and Moritz Münchmeyer. "Normalizing flows for random fields in cosmology". In: *arXiv e-prints*, arXiv:2105.12024 (May 2021), arXiv:2105.12024. arXiv: 2105.12024 [astro-ph.CO].

[275] Douglas H Rudd, Andrew R Zentner, and Andrey V Kravtsov. "Effects of baryons and dissipation on the matter power spectrum". In: *The Astrophysical Journal* 672.1 (2008), p. 19.

[276] Jun'ichi Sato et al. "Implication of Ωm through the morphological analysis of weak lensing fields". In: *The Astrophysical Journal* 551.1 (2001), p. L5.

[277] Anna M. M. Scaife and Fiona Porter. "Fanaroff-Riley classification of radio galaxies using group-equivariant convolutional neural networks". In: *Monthly Notices of the Royal Astronomical Society* 503.2 (May 2021), pp. 2369–2379. DOI: 10.1093/mnras/stab530. arXiv: 2102.08252 [astro-ph.IM].

[278] Joop Schaye et al. "The EAGLE project: simulating the evolution and assembly of galaxies and their environments". In: *MNRAS* 446.1 (Jan. 2015), pp. 521–554. DOI: 10.1093/mnras/stu2058. arXiv: 1407.7040 [astro-ph.GA].

[279] Joop Schaye et al. "The physics driving the cosmic star formation history". In: *MNRAS* 402.3 (Mar. 2010), pp. 1536–1560. DOI: 10.1111/j.1365-2966.2009.16029.x. arXiv: 0909.5196 [astro-ph.CO].

[280] Fabian Schmidt. "Weak lensing probes of modified gravity". In: *Phys. Rev. D* 78.4, 043002 (Aug. 2008), p. 043002. DOI: 10.1103/PhysRevD.78.043002. arXiv: 0805.4812 [astro-ph].

[281] A. Schneider and R. Teyssier. "A new method to quantify the effects of baryons on the matter power spectrum". In: *J. Cosmology Astropart. Phys.* 12, 049 (Dec. 2015), p. 049. DOI: 10.1088/1475-7516/2015/12/049. arXiv: 1510.06034.

[282] Aurel Schneider et al. "Baryonic effects for weak lensing. Part II. Combination with X-ray data and extended cosmologies". In: *Journal of Cosmology and Astroparticle Physics* 2020.04 (Apr. 2020), pp. 020–020. ISSN: 1475-7516. DOI: 10.1088/1475-7516/2020/04/020. URL: http://dx.doi.org/10.1088/1475-7516/2020/04/020.

[283] Aurel Schneider et al. "Quantifying baryon effects on the matter power spectrum and the weak lensing shear correlation". In: *Journal of Cosmology and Astroparticle Physics* 2019.03 (Mar. 2019), pp. 020–020. ISSN: 1475-7516. DOI: 10.1088/1475-7516/2019/03/020. URL: http://dx.doi.org/10.1088/1475-7516/2019/03/020.

[284] P. Schneider, M. Kilbinger, and M. Lombardi. "The three-point correlation function of cosmic shear: II. Relation to the bispectrum of the projected mass density and generalized third-order aperture measures". In: *Astronomy & Astrophysics* 431.1 (Feb. 2005), pp. 9–25. ISSN: 1432-0746. DOI: 10.1051/0004-6361:20034217. URL: http://dx.doi.org/10.1051/0004-6361:20034217.

[285] Peter Schneider, Jürgen Ehlers, and Emilio E. Falco. *Gravitational Lenses*. 1992. DOI: 10.1007/978-3-662-03758-4.

[286] Emiliano Sefusatti et al. "Cosmology and the Bispectrum". In: *Physical Review D* 74.2 (2006), p. 023522.

[287] U. Seljak et al. "Towards optimal extraction of cosmological information from nonlinear data". In: *J. Cosmology Astropart. Phys.* 12, 009 (Dec. 2017), p. 009. DOI: 10.1088/1475-7516/2017/12/009. arXiv: 1706.06645.

[288] Uroš Seljak. "Cosmography and Power Spectrum Estimation: A Unified Approach". In: *ApJ* 503.2 (Aug. 1998), pp. 492–501. DOI: 10.1086/306019. arXiv: astro-ph/9710269 [astro-ph].

[289] Uros Seljak and Matias Zaldarriaga. "A Line-of-Sight Integration Approach to Cosmic Microwave Background Anisotropies". In: *ApJ* 469 (Oct. 1996), p. 437. DOI: 10.1086/177793. arXiv: astro-ph/9603033 [astro-ph].

[290] Elisabetta Semboloni, Tim Schrabback, and Ludovic van Waerbeke et al. "Weak lensing from space: first cosmological constraints from three-point shear statistics". In: *MNRAS* 410.1 (Jan. 2011), pp. 143–160. DOI: 10.1111/j.1365-2966.2010.17430.x. arXiv: 1005.4941 [astro-ph.CO].

[291] Divij Sharma, Biwei Dai, and Uros Seljak. "A comparative study of cosmological constraints from weak lensing using Convolutional Neural Networks". In: *arXiv e-prints*, arXiv:2403.03490 (Mar. 2024), arXiv:2403.03490. DOI: 10.48550/arXiv.2403.03490. arXiv: 2403.03490 [astro-ph.CO].

[292] Divij Sharma et al. "A field-level emulator for modeling baryonic effects across hydrodynamic simulations". In: *arXiv e-prints*, arXiv:2401.15891 (Jan. 2024), arXiv:2401.15891. DOI: 10.48550/arXiv.2401.15891. arXiv: 2401.15891 [astro-ph.CO].

[293] Théo Simon et al. "Consistency of effective field theory analyses of the BOSS power spectrum". In: *Phys. Rev. D* 107.12, 123530 (June 2023), p. 123530. DOI: 10.1103/PhysRevD.107.123530. arXiv: 2208.05929 [astro-ph.CO].

[294] Michael J Smith et al. "Realistic galaxy image simulation via score-based generative models". In: *Monthly Notices of the Royal Astronomical Society* 511.2 (2022), pp. 1808–1818.

[295] Robert E Smith et al. "Precision cosmology in muddy waters: cosmological constraints and N-body codes". In: *Monthly Notices of the Royal Astronomical Society* 440.1 (2014), pp. 249–268.

[296] Yang Song and Diederik P Kingma. "How to train your energy-based models". In: *arXiv preprint arXiv:2101.03288* (2021).

[297] Yong-Seon Song and Lloyd Knox. "Determination of cosmological parameters from cosmic shear data". In: *Phys. Rev. D* 70.6, 063510 (Sept. 2004), p. 063510. DOI: 10.1103/PhysRevD.70.063510. arXiv: astro-ph/0312175 [astro-ph].

[298] D. Spergel, N. Gehrels, and C. Baltay et al. "Wide-Field InfrarRed Survey Telescope-Astrophysics Focused Telescope Assets WFIRST-AFTA 2015 Report". In: *arXiv e-prints*, arXiv:1503.03757 (Mar. 2015), arXiv:1503.03757. arXiv: 1503.03757 [astro-ph.IM].

[299] V. Springel. "E pur si muove: Galilean-invariant cosmological hydrodynamical simulations on a moving mesh". In: *MNRAS* 401 (Jan. 2010), pp. 791–851. DOI: 10.1111/j.1365-2966.2009.15715.x. arXiv: 0901.4107.

[300] V. Springel et al. "First results from the IllustrisTNG simulations: matter and galaxy clustering". In: *MNRAS* 475 (Mar. 2018), pp. 676–698. DOI: 10.1093/mnras/stx3304. arXiv: 1707.03397.

[301] Volker Springel. "The cosmological simulation code GADGET-2". In: *Monthly Notices of the Royal Astronomical Society* 364.4 (Dec. 2005), pp. 1105–1134. DOI: 10.1111/j.1365-2966.2005.09655.x. arXiv: astro-ph/0505010 [astro-ph].

[302] Volker Springel et al. “First results from the IllustrisTNG simulations: matter and galaxy clustering”. In: *Monthly Notices of the Royal Astronomical Society* 475.1 (2018), pp. 676–698.

[303] Volker Springel et al. “Simulations of the formation, evolution and clustering of galaxies and quasars”. In: *nature* 435.7042 (2005), pp. 629–636.

[304] J-L Starck and Fionn Murtagh. *Astronomical image and data analysis*. Springer Science & Business Media, 2007.

[305] Rainer Storn. “On the usage of differential evolution for function optimization”. In: July 1996, pp. 519–523. ISBN: 0-7803-3225-3. DOI: 10.1109/NAFIPS.1996.534789.

[306] Rainer Storn and Kenneth Price. “Differential Evolution - A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces”. In: *Journal of Global Optimization* 11 (Jan. 1997), pp. 341–359. DOI: 10.1023/A:1008202821328.

[307] T. Sunayama et al. “Efficient construction of mock catalogs for baryon acoustic oscillation surveys”. In: *J. Cosmology Astropart. Phys.* 5, 051 (May 2016), p. 051. DOI: 10.1088/1475-7516/2016/05/051. arXiv: 1510.06665.

[308] Masahiro Takada and Bhuvnesh Jain. “The kurtosis of the cosmic shear field”. In: *Monthly Notices of the Royal Astronomical Society* 337.3 (2002), pp. 875–894.

[309] Masahiro Takada and Bhuvnesh Jain. “Three-point correlations in weak lensing surveys: model predictions and applications”. In: *MNRAS* 344.3 (Sept. 2003), pp. 857–886. DOI: 10.1046/j.1365-8711.2003.06868.x. arXiv: astro-ph/0304034 [astro-ph].

[310] R. Takahashi et al. “Revising the Halofit Model for the Nonlinear Matter Power Spectrum”. In: *ApJ* 761, 152 (Dec. 2012), p. 152. DOI: 10.1088/0004-637X/761/2/152. arXiv: 1208.2701.

[311] Atsushi Taruya et al. “Lognormal property of weak-lensing fields”. In: *The Astrophysical Journal* 571.2 (2002), p. 638.

[312] S. Tassev, M. Zaldarriaga, and D. J. Eisenstein. “Solving large scale structure in ten easy steps with COLA”. In: *J. Cosmology Astropart. Phys.* 6, 036 (June 2013), p. 036. DOI: 10.1088/1475-7516/2013/06/036. arXiv: 1301.0322 [astro-ph.CO].

[313] Peter L Taylor, Francis Bernardeau, and Thomas D Kitching. “k-cut cosmic shear: Tunable power spectrum sensitivity to test gravity”. In: *Physical Review D* 98.8 (2018), p. 083514.

[314] Peter L. Taylor, Thomas D. Kitching, and Justing Alsing et al. “Cosmic shear: Inference from forward models”. In: *Phys. Rev. D* 100.2, 023519 (July 2019), p. 023519. DOI: 10.1103/PhysRevD.100.023519. arXiv: 1904.05364 [astro-ph.CO].

[315] Max Tegmark. “How to measure CMB power spectra without losing information”. In: *Physical Review D* 55.10 (May 1997), pp. 5895–5907. DOI: 10.1103/PhysRevD.55.5895. arXiv: astro-ph/9611174 [astro-ph].

[316] R. Teyssier. "Cosmological hydrodynamics with adaptive mesh refinement. A new high resolution code called RAMSES". In: *A&A* 385 (Apr. 2002), pp. 337–364. DOI: 10.1051/0004-6361:20011817. arXiv: astro-ph/0111367 [astro-ph].

[317] R. Teyssier et al. "Mass distribution in galaxy clusters: the role of Active Galactic Nuclei feedback". In: *MNRAS* 414 (June 2011), pp. 195–208. DOI: 10.1111/j.1365-2966.2011.18399.x. arXiv: 1003.4744.

[318] Lucas Theis, Aäron van den Oord, and Matthias Bethge. "A note on the evaluation of generative models". In: *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings*. Ed. by Yoshua Bengio and Yann LeCun. 2016. URL: http://arxiv.org/abs/1511.01844.

[319] Tijmen Tieleman. "Training restricted Boltzmann machines using approximations to the likelihood gradient". In: *Proceedings of the 25th international conference on Machine learning*. 2008, pp. 1064–1071.

[320] P. Torrey et al. "A model for cosmological simulations of galaxy formation physics: multi-epoch validation". In: *MNRAS* 438 (Mar. 2014), pp. 1985–2004. DOI: 10.1093/mnras/stt2295. arXiv: 1305.4931.

[321] Tilman Tröster et al. "Painting with baryons: augmenting N-body simulations with gas using deep generative models". In: *Monthly Notices of the Royal Astronomical Society: Letters* 487.1 (2019), pp. L24–L29.

[322] A. Vale and J. P. Ostriker. "Linking halo mass to galaxy luminosity". In: *MNRAS* 353.1 (Sept. 2004), pp. 189–200. DOI: 10.1111/j.1365-2966.2004.08059.x. arXiv: astro-ph/0402500 [astro-ph].

[323] Georgios Valogiannis and Cora Dvorkin. "Going beyond the galaxy power spectrum: An analysis of BOSS data with wavelet scattering transforms". In: *Physical Review D* 106.10, 103509 (Nov. 2022), p. 103509. DOI: 10.1103/PhysRevD.106.103509. arXiv: 2204.13717 [astro-ph.CO].

[324] Georgios Valogiannis and Cora Dvorkin. "Towards an optimal estimation of cosmological parameters with the wavelet scattering transform". In: *Physical Review D* 105.10 (May 2022). ISSN: 2470-0029. DOI: 10.1103/physrevd.105.103534. URL: http://dx.doi.org/10.1103/PhysRevD.105.103534.

[325] M. P. van Daalen and J. Schaye. "The contributions of matter inside and outside of haloes to the matter power spectrum". In: *MNRAS* 452 (Sept. 2015), pp. 2247–2257. DOI: 10.1093/mnras/stv1456. arXiv: 1501.05950.

[326] M. P. van Daalen et al. "The effects of galaxy formation on the matter power spectrum: a challenge for precision cosmology". In: *MNRAS* 415 (Aug. 2011), pp. 3649–3665. DOI: 10.1111/j.1365-2966.2011.18981.x. arXiv: 1104.1174 [astro-ph.CO].

[327] Marcel P. van Daalen, Ian G. McCarthy, and Joop Schaye. "Exploring the effects of galaxy formation on matter clustering through a library of simulation power spectra". In: *MNRAS* 491.2 (Jan. 2020), pp. 2424–2446. DOI: 10.1093/mnras/stz3199. arXiv: 1906.00968 [astro-ph.CO].

[328] M. Velliscig et al. "The impact of galaxy formation on the total mass, mass profile and abundance of haloes". In: *MNRAS* 442 (Aug. 2014), pp. 2641–2658. DOI: 10.1093/mnras/stu1044. arXiv: 1402.4461.

[329] M. Vicinanza et al. "Higher order moments of lensing convergence - I. Estimate from simulations". In: *arXiv e-prints*, arXiv:1606.03892 (June 2016), arXiv:1606.03892. DOI: 10.48550/arXiv.1606.03892. arXiv: 1606.03892 [astro-ph.CO].

[330] Martina Vicinanza et al. "Increasing the lensing figure of merit through higher order convergence moments". In: *Physical Review D* 97.2, 023519 (Jan. 2018), p. 023519. DOI: 10.1103/PhysRevD.97.023519. arXiv: 1802.02963 [astro-ph.CO].

[331] Martina Vicinanza et al. "Minkowski functionals of convergence maps and the lensing figure of merit". In: *Physical Review D* 99.4, 043534 (Feb. 2019), p. 043534. DOI: 10.1103/PhysRevD.99.043534. arXiv: 1905.00410 [astro-ph.CO].

[332] Ricardo Vilalta. *Transfer Learning in Astronomy: A New Machine-Learning Paradigm*. 2018. arXiv: 1812.10403 [astro-ph.HE].

[333] Francisco Villaescusa-Navarro et al. "Multifield Cosmology with Artificial Intelligence". In: *arXiv preprint arXiv:2109.09747* (2021).

[334] Francisco Villaescusa-Navarro et al. *Neural networks as optimal estimators to marginalize over baryonic effects*. 2020. arXiv: 2011.05992 [astro-ph.CO].

[335] Francisco Villaescusa-Navarro et al. "Robust marginalization of baryonic effects for cosmological inference at the field level". In: *arXiv e-prints*, arXiv: 2109.10360 (Sept. 2021). DOI: 10.48550/arXiv.2109.10360. arXiv: 2109.10360 [astro-ph.CO].

[336] Francisco Villaescusa-Navarro et al. "The CAMELS Multifield Dataset: Learning the Universe's Fundamental Parameters with Artificial Intelligence". In: *arXiv e-prints*, arXiv: 2109.10915 (Sept. 2021), arXiv: 2109.10915. arXiv: 2109.10915 [cs.LG].

[337] Francisco Villaescusa-Navarro et al. "The CAMELS Project: Cosmology and Astrophysics with Machine-learning Simulations". In: *ApJ* 915.1, 71 (July 2021), p. 71. DOI: 10.3847/1538-4357/abf7ba. arXiv: 2010.00619 [astro-ph.CO].

[338] Francisco Villaescusa-Navarro et al. "The CAMELS Project: Public Data Release". In: *The Astrophysical Journal Supplement Series* 265.2 (Apr. 2023), p. 54. ISSN: 1538-4365. DOI: 10.3847/1538-4365/acbf47. URL: http://dx.doi.org/10.3847/1538-4365/acbf47.

[339] Pablo Villanueva-Domingo and Francisco Villaescusa-Navarro. "Learning cosmology and clustering with cosmic graphs". In: *The Astrophysical Journal* 937.2 (2022), p. 115.

[340] Pauli Virtanen et al. "SciPy 1.0: fundamental algorithms for scientific computing in Python". In: *Nature Methods* 17.3 (Feb. 2020), pp. 261–272. ISSN: 1548-7105. DOI: 10.1038/s41592-019-0686-2. URL: http://dx.doi.org/10.1038/s41592-019-0686-2.

[341] M. Vogelsberger et al. "Introducing the Illustris Project: simulating the coevolution of dark and visible matter in the Universe". In: *MNRAS* 444 (Oct. 2014), pp. 1518–1547. DOI: 10.1093/mnras/stu1536. arXiv: 1405.2921.

[342] M. Vogelsberger et al. "Properties of galaxies reproduced by a hydrodynamic simulation". In: *Nature* 509 (May 2014), pp. 177–182. DOI: 10.1038/nature13316. arXiv: 1405.1418.

[343] Mark Vogelsberger et al. "A model for cosmological simulations of galaxy formation physics". In: *MNRAS* 436.4 (Dec. 2013), pp. 3031–3067. DOI: 10.1093/mnras/stt1789. arXiv: 1305.2913 [astro-ph.CO].

[344] Cunshi Wang et al. *Transfer Learning Applied to Stellar Light Curve Classification*. 2023. arXiv: 2305.13745 [astro-ph.IM].

[345] H. Wang et al. "Reconstructing the Initial Density Field of the Local Universe: Methods and Tests with Mock Catalogs". In: *ApJ* 772, 63 (July 2013), p. 63. DOI: 10.1088/0004-637X/772/1/63. arXiv: 1301.1348.

[346] Huiyuan Wang et al. "ELUCID—Exploring the Local Universe with the Reconstructed Initial Density Field. I. Hamiltonian Markov Chain Monte Carlo Method with Particle Mesh Dynamics". In: *The Astrophysical Journal* 794.1, 94 (Oct. 2014), p. 94. DOI: 10.1088/0004-637X/794/1/94. arXiv: 1407.3451 [astro-ph.CO].

[347] Rui Wang, Robin Walters, and Rose Yu. "Incorporating Symmetry into Deep Dynamics Models for Improved Generalization". In: *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net, 2021. URL: https://openreview.net/forum?id=wta%5C_8Hx2KD.

[348] Maurice Weiler and Gabriele Cesa. "General E(2)-Equivariant Steerable CNNs". In: *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*. Ed. by Hanna M. Wallach et al. 2019, pp. 14334–14345.

[349] Maurice Weiler et al. "3D Steerable CNNs: Learning Rotationally Equivariant Features in Volumetric Data". In: *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montréal, Canada*. Ed. by Samy Bengio et al. 2018, pp. 10402–10413.

[350] David H. Weinberg. "Reconstructing primordial density fluctuations. I - Method". In: *Monthly Notices of the Royal Astronomical Society* 254 (Jan. 1992), pp. 315–342. DOI: 10.1093/mnras/254.2.315.

[351] Rainer Weinberger, Volker Springel, and Rüdiger Pakmor. "The AREPO Public Code Release". In: *ApJs* 248.2, 32 (June 2020), p. 32. DOI: 10.3847/1538-4365/ab908c. arXiv: 1909.04667 [astro-ph.IM].

[352] Martin White. "A marked correlation function for constraining modified gravity models". In: *Journal of Cosmology and Astroparticle Physics* 2016.11 (2016), p. 057.

[353] Simon DM White. "The hierarchy of correlation functions and its relation to other measures of galaxy clustering". In: *Monthly Notices of the Royal Astronomical Society* 186.2 (1979), pp. 145–154.

[354] Hans A. Winther et al. "Emulators for the nonlinear matter power spectrum beyond ΛCDM". In: *Physical Review D* 100.12 (Dec. 2019). ISSN: 2470-0029. DOI: 10.1103/physrevd.100.123540. URL: http://dx.doi.org/10.1103/PhysRevD.100.123540.

[355] Daniel E Worrall et al. "Harmonic networks: Deep translation and rotation equivariance". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017, pp. 5028–5037.

[356] Henrique S Xavier, Filipe B Abdalla, and Benjamin Joachimi. "Improving lognormal models for cosmological fields". In: *Monthly Notices of the Royal Astronomical Society* 459.4 (2016), pp. 3693–3710.

[357] Xiuyuan Yang et al. "Cosmological information in weak lensing peaks". In: *Physical Review D* 84.4 (2011), p. 043529.

[358] Dani Yogatama et al. "Generative and Discriminative Text Classification with Recurrent Neural Networks". In: *CoRR* abs/1703.01898 (2017). arXiv: 1703.01898. URL: http://arxiv.org/abs/1703.01898.

[359] Jason J Yu, Konstantinos G Derpanis, and Marcus A Brubaker. "Wavelet flow: Fast training of high resolution normalizing flows". In: *Advances in Neural Information Processing Systems* 33 (2020), pp. 6184–6196.

[360] Matias Zaldarriaga and Roman Scoccimarro. "Higher Order Moments of the Cosmic Shear and Other Spin-2 Fields". In: *The Astrophysical Journal* 584.2 (Feb. 2003), pp. 559–565. ISSN: 1538-4357. DOI: 10.1086/345789. URL: http://dx.doi.org/10.1086/345789.

[361] Justine Zeghal et al. *Neural Posterior Estimation with Differentiable Simulators*. 2022. arXiv: 2207.05636 [astro-ph.IM].

[362] Ya. B. Zel'dovich. "Gravitational instability: An approximate theory for large density perturbations." In: *A&A* 5 (Mar. 1970), pp. 84–89.

[363] Zhongxu Zhai et al. "The Aemulus Project. III. Emulation of the Galaxy Correlation Function". In: *ApJ* 874.1, 95 (Mar. 2019), p. 95. DOI: 10.3847/1538-4357/ab0d7b. arXiv: 1804.05867 [astro-ph.CO].

[364] Xinyue Zhang et al. "From Dark Matter to Galaxies with Convolutional Networks". In: *arXiv preprint arXiv:1902.05965* 2019 (2019).

[365] Xiaosheng Zhao et al. "Can Diffusion Model Conditionally Generate Astrophysical Images?" In: *arXiv preprint arXiv:2307.09568* (2023).

[366] Chenyu Zheng et al. "Revisiting Discriminative vs. Generative Classifiers: Theory and Implications". In: *International Conference on Machine Learning, ICML 2023, 23-29 July 2023, Honolulu, Hawaii, USA*. Ed. by Andreas Krause et al. Vol. 202. Proceedings of Machine Learning Research. PMLR, 2023, pp. 42420–42477. URL: https://proceedings.mlr.press/v202/zheng23f.html.

[367] Fuzhen Zhuang et al. "A comprehensive survey on transfer learning". In: *Proceedings of the IEEE* 109.1 (2020), pp. 43–76.

[368] A. Zonca et al. "healpy: equal area pixelization and spherical harmonics transforms for data on the sphere in Python". In: *The Journal of Open Source Software* 4 (Mar. 2019), p. 1298. DOI: 10.21105/joss.01298.

[369] D. Zürcher et al. "Towards a full wCDM map-based analysis for weak lensing surveys". In: *Monthly Notices of the Royal Astronomical Society* 525.1 (Oct. 2023), pp. 761–784. DOI: 10.1093/mnras/stad2212. arXiv: 2206.01450 [astro-ph.CO].

[370] Dominik Zürcher et al. "Dark energy survey year 3 results: Cosmology with peaks using an emulator approach". In: *Monthly Notices of the Royal Astronomical Society* 511.2 (2022), pp. 2075–2104.