UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title

Modern Statistical Methods for Large-scale Structure Cosmology

Permalink

Author

Sullivan, James Michael

Publication Date

2024

Peer reviewed|Thesis/dissertation

Modern Statistical Methods for Large-scale Structure Cosmology

by

James M. Sullivan

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Astrophysics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Uroš Seljak, Chair
Professor Martin White
Professor Joshua S. Bloom

Summer 2024

Modern Statistical Methods for Large-scale Structure Cosmology

Abstract

Modern Statistical Methods for Large-scale Structure Cosmology

by

James M. Sullivan

Doctor of Philosophy in Astrophysics

University of California, Berkeley

Professor Uroš Seljak, Chair

The number and quality of large-scale structure (LSS) surveys is already stretching current methods of learning about cosmology from data to their capacity. However, the pace of data streaming in will increase significantly, with, at the time of writing, data from the Dark Energy Spectroscopic Instrument (DESI), Euclid, Rubin, Spectro-Photometer for the History of the Universe, Epoch of Reionization and Ices Explorer (SPHEREx), Spec-S5, and Roman joining existing data from the Baryon Oscillation Spectroscopic Survey (BOSS), Dark Energy Survey (DES), and Kilo Degree Survey (KiDS) as well as CMB lensing data. Maximally extracting cosmological information from these datasets depends, of course, on high-fidelity understanding of the instruments and observational effects involved in generating the data. But more fundamentally, the cosmological information accessed with these surveys is generated by non-linear, non-perturbative, and high-dynamic range physical processes.

In the face of highly-constraining data, accurate and precise models for such processes are necessarily complex. This generates several challenges for using them to access cosmological information. However, recent advances in high-performance computing have enabled more massive and high-resolution cosmological numerical simulations than ever before. Simultaneously, the growth of Graphics Processing Units (GPUs) and related tooling, such as methods of automatic differentiation, have led to an explosion of machine learning architectures and methods for working with high dimensional models and data, with concomitant application of said methods to cosmological problems by the LSS community. These computational advances have been, and will continue to be, drafted into the service of extracting cosmological information from high-quality data. This thesis highlights several areas where recent technological developments can be directly translated into improved methods for large-scale structure simulation and analysis.

Increasingly complex models generate additional parameters, which, though typically not of cosmological interest, must be included and varied - leading to a more challenging inference problem and, frequently, to less interpretable phenomenological models of LSS. One machine-

learning-informed strategy to address this issue in the context of simulation-based prior assumptions is outlined in Chapter 2, while a more direct strategy for speeding up the initial phase of the inference procedure using machine learning methods in a more general inference context is the subject of Chapter 5.

More straightforwardly, raw computational cost also grows when numerical models are asked to describe a wider range of scales, which will be necessary for high-density LSS tracer samples covering large spatial volumes and redshift ranges. Chapter 3 details a scheme for improving numerical simulation efficiency, therefore reducing this growing computational cost in the context of modeling the cosmological impact of massive neutrinos on LSS. Numerical simulations that attempt to model galaxy formation in a cosmological context are also increasingly being used to inform LSS tracer properties. As these simulations become more robust in their determination of tracer population properties, strategies for leveraging these properties will enhance accessible cosmological information from surveys. An example for performing such leveraging with a machine learning-based strategy in the context of primordial non-Gaussianity is outlined in Chapter 4. Extending methods in similar directions going forward will enable the LSS community to learn the most possible from large-scale structure survey data.

To my parents, who made it possible for me to choose this path.

Contents

List of Figures

List of Tables

Acknowledgments

Let me start this incomplete list of thank-yous with those who, early on, influenced my eventual decision to enter a PhD program. First, I thank Lee Ann Hennig for truly dedicating herself to communicating the awe and magic of astronomy and astrophysics at TJ, as well as nurturing the naive dreams of a student who didn't exactly excel in physics and math courses at the time. I also want to thank Drs. Shing Fung and Leonard Garcia, who prodded me through my first "real" physics project mapping out the beaming geometry of auroral kilometric radiation from spacecraft data, the thrill of which sent me to Texas to pursue physics and astronomy (and math). I want to express gratitude for the Texas mentorship of Profs. J. Craig Wheeler, Volker Bromm, and Dr. Mike Montgomery, who taught me not only about supernovae, the first stars, and the "last stars" (white dwarfs), but more importantly the elements of astrophysics research, specifically of the computational variety. I also thank Dr. Alexander Wiegand and Prof. Daniel Eisenstein, who invited me into the world of large-scale structure cosmology and emphasized the importance of statistics, as well as Drs. Jonathan McDowell and Matthew Ashby, who made that invitation possible.

At Berkeley, first, I of course have to thank my PhD advisor Uroš. By way of his incredibly sharp insights, relentless shucking of any problem to its core, and career guidance through many episodes, I have learned to do research more and more effectively. Thanks are also in order for demonstrating immense patience with me as well as allowing me to chase after my own interests, despite the delays incurred, and, most of all, for treating me as a scientific adult, even (especially) when I was anything but. I thank Martin for his incredibly wide-ranging scientific advice and constructive comments throughout my time at Berkeley, for demonstrating how to zero in on bedrock physical principles, as well as for thoughtful support when planning for the future. Thanks also to Salman, for intelligent, warm, and candid guidance as well as to Katrin and the CPAC group for being so welcoming and setting a wonderful example for how a large, bleeding-edge computational cosmology group can be run. Also, thanks to JD Emberson and Matt Becker for seriously and positively mentoring me at Argonne. Thank you also to Prof. Josh Bloom for serving on all of my committees at Berkeley and for creative instruction in high-energy astrophysics.

Thanks to all the grads in the BCCP - Stephen for many long and stimulating conversations about life and, of course, science - Adrian for many fun times had and constructive doubts - Byeonghee for being a patient early-days resource even when most of your help refused to stick and slid immediately onto the floor - Biwei for camaraderie and spreading confusion - and all of the other BCCP students, postdocs, and senior members - especially Sukhdeep for invaluable early mentorship, as well as Antón, Minas, Joanne, Joe, Noah, Noah, Yu, Alex, and Boryana, Roger, Simone, and Pat, for scientific wisdom and discussions ranging far and wide. A special thanks goes to Zack for energizing collaboration and Marius, Marco, and Jaime for Julia expertise and enthusiasm. I also have to mention the BADgrads, especially Nick, Aliza, Massimo, Emma, and Isaac (honorary) for 3rd-floor discussions plus related activities, Andrea, Kishore and the rest of my cohort for mutual support teaching and taking the astro courses, and all for tying together a strong community. Thanks also

Chapter 1

Introduction

First, a disclaimer. Here we will only have space to give a quick overview of modern large-scale structure cosmology, specifically focusing on models. The hope is that this will give the nonspecialized reader some needed context should they choose to move on to subsequent chapters. This necessarily means disregarding many works - this introduction will be neither complete nor self-contained - some review articles and relevant texts will be referenced where omissions become particularly glaring. However, we will make an effort to string relevant facts together in a fashion that gives some peek at the rich tapestry of cosmological physics. We will use the notation that $\dot{} = \frac{d}{dt}$, $' = \frac{d}{d\tau}$, $c = \hbar = 1$, and the (-,+,+,+) metric convention.

Cosmology is chiefly concerned with several aspects of the universe - its global geometry, its constituents, and how they move on large scales. All of these aspects are well-described by the standard Λ cold dark matter (ΛCDM) model, along with a couple of extensions. Stated another way, the structure of the standard cosmological model reduces to solving the Einstein equations

$$G_{\mu\nu} = 8\pi G T_{\mu\nu}, \tag{1.1}$$

under several symmetry assumptions. The evolution of both the energy density of the universe as a whole and its large-scale fluctuations is governed by a handful of equations that take only 10 numbers as inputs:

1. Ω_k - the global curvature,

2. h - the expansion rate (Hubble parameter),

3. Ω_c - the relative amount of cold dark matter,

4. Ω_b - the relative amount of baryonic (non-dark) matter,

5. Ω_γ - the relative amount of photons,

6. Ω_Λ - the relative amount of dark energy[1],

---

[1] We will not discuss evolving dark energy here, but see Ch. 28 of Ref. [1] as a starting place.

7. N_{eff} - the amount of light species,

8. $\sum m_\nu$ - the sum of neutrino masses[2],

9. A_s - the amplitude of primordial fluctuations,

10. n_s - the deviation from scale-independence of primordial fluctuations,

where we will further discuss each of these parameters below. Recently determined numerical values for these parameters can be found in Ref. [3]. The point to stress here is that all numbers above are generally known to the several percent level or better from the combination of several cosmological observations. The rest of this Chapter aims to introduce the reader to the handful of equations necessary to tell the story of how the universe at large evolves from its earliest moments to today. We will start with the global evolution of the energy density.

Following from spatial translation and rotation invariance in general relativity, we obtain the Friedmann-Lemaître-Robinson-Walker (FLRW) metric

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

where k encodes the global curvature. When k is positive, a fixed-time hypersurface exhibits spherical geometry; when zero, flat geometry; and when negative, hyperbolic geometry. This number can be related to Ω_k as $\Omega_k = -\left(\frac{k}{H_0} \right)^2$, where $H_0 = 100 \frac{\text{km}}{s \cdot \text{Mpc}} h$. Since Ω_k is constrained by data to be very close to zero, for the remainder of this dissertation we will work under the often-used assumption of flat global geometry, $\Omega_k = 0$.

Given this metric, at the level of the background (i.e., global, or spatially-averaged on a given time slice) the Einstein equations 1.1 reduce to the flat-space Friedmann equations

$$\left(\frac{\dot{a}}{a} \right)^2 = \frac{8\pi G}{3} \rho \qquad (1.3)$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3} \left(\rho + 3P \right), \qquad (1.4)$$

where a is the scale factor (set to 1 today), ρ, P are the global density and pressure as defined by $T^\nu_\mu = \text{diag}(-\rho, P, P, P)$.

We can rewrite the density in eqn. 1.3 as $\rho = \sum_X \Omega_X \rho_{\text{crit}}$ where $\rho_{\text{crit}} = \frac{3H_0^2}{8\pi G}$ and $\Omega_X = \frac{\rho_X}{\rho_{\text{crit}}}$. We can now write the Hubble function $H(a) = \frac{\dot{a}}{a}$ that describes the expansion history of the universe by rewriting the first Friedmann equation

$$\left(\frac{H(a)}{H_0} \right)^2 = \Omega_\gamma \left(1 + \frac{7}{8} \left(\frac{4}{11} \right)^{\frac{4}{3}} N_{\text{eff}} \right) a^{-4} + \left(\Omega_c + \Omega_b \right) a^{-3} + \Omega_{m\nu}(a; \sum m_\nu) + \Omega_\Lambda, \quad (1.5)$$

---

[2]Cosmological perturbations are sensitive to more detailed neutrino physics, but this is the most widely-considered parameter [2].

which exposes the first 8 of our 10 parameters enumerated above ($\Omega_{m\nu}$ is a non-power-law function of a that depends on $\sum m_\nu$). From the power law scalings and known numerical values of the Ω_X parameters, equation 1.5 indicates that in turns, the energy density of the universe will be dominated first by radiation (photons and neutrinos, or other light species), then by matter (CDM and baryons), and finally by dark energy (here modeled as a constant contribution, Λ).

This concludes our brief overview of the spatially-averaged dynamics of the universe[3]. The story obviously does not end there, as there is abundant structure in the universe in every direction we look - galaxies, galaxy clusters, the circum/intergalactic medium, huge filaments and walls of galaxies, and even anisotropic microwaves being emitted from all directions. Where does this structure in the universe arise from?

To address this question, we must turn our attention to perturbations in the fields of interest - spatial fluctuations in the metric and the matter/radiation fluctuations that we treat as classical random fields deviating from a smooth background. We also need a mechanism for generating large-scale homogeneity, isotropy, and flatness in the first place. The next section will touch on inflation, a mechanism for generating both large-scale homogeneity and isotropy, observed flatness, as well as spatial fluctuations to seed the large-scale structure we observe. The salient details will also introduce us to the last two parameters above, A_s and n_s.

1.1 Initial conditions

Here we briefly review the elements of inflation relevant for the standard model of cosmology. For more details see, e.g., Refs. [4, 5].

Given some initial conditions, inflation is at its most basic a mechanism to generate a largely flat, homogeneous, and isotropic patch big enough to be consistent with the observed universe, which demonstrates all these qualities on scales larger than roughly 100 Mpc. Though usually stated in terms of observational puzzles (see Figure 1.1), including the horizon problem (the universe appears uniform on length scales even when causal connection is impossible) and flatness problem (the extremely small value of observed global curvature), the origin of large-scale homogeneity and isotropy can also be phrased in terms of initial conditions of the density and velocity field. Said another way, the homogeneity and isotropy of the universe suggests finely-tuned initial conditions for the energy density, while flatness suggests the same for initial conditions of the velocity field [5]. Inflation acts to explain how an initially small and largely homogeneous and isotropic region can become large.

Inflation also *generates* the perturbations that eventually lead to the nonlinear structure we see in LSS on smaller scales today. To quickly see this, we can consider the usual case of a scalar field $\varphi = \varphi(\mathbf{x}, \tau)$ as a model for inflation[4]. For a classical scalar field with arbitrary

---

[3]We won't discuss Big Bang Nucleosynthesis (BBN) or the thermal history of the universe, but see Section 3 here for an overview.
45

potential, we write the action as

$$S = \int d^4x \sqrt{-g} \left(-\frac{1}{2} g^{\mu\nu} \partial_\mu \varphi \partial_\nu \varphi \right) \tag{1.6}$$

Varying this action leads to the Klein-Gordon equation of motion, which, when splitting the field as $\varphi = \bar{\varphi} + \delta\varphi$, leads to, in spatially-flat gauge, the Mukhanov-Sasaki equation

$$f'' = - \left(k^2 - \frac{z''}{z} \right) f, \tag{1.7}$$

where $z = \frac{\bar{\varphi}'}{H}$ and $f = a\delta\varphi$. In terms of these variables, we can write the quadratic action, which can be used to quantize the theory. Following the quantization process and the assumption of a particular vacuum state, we can solve for the mode function solutions $f(\mathbf{k}, \tau)$ (for a particular Fourier mode $\mathbf{k}$) of this equation that describe the perturbations.

What does the structure of these perturbations look like? Rather than answer this question for f, for large-scale structure we are most interested in a related quantity, the comoving curvature perturbation $\mathcal{R} = -\frac{f}{z} = \frac{-\delta\varphi H}{\bar{\varphi}'}$. $\mathcal{R}$ is closely connected to observables, as it is constant at linear order on very large scales - i.e. for modes with wavelengths larger than the horizon scale $(aH)^{-1}$, and therefore its superhorizon value set by inflation serves as the initial condition for any particular Fourier mode of cosmological interest. For Gaussian fluctuations arising from the simple scalar field quantization, the fluctuations are completely described by their power spectrum[6].

At the end of the day, when the Mukhanov-Sasaki equation is solved (numerically or through approximations), one obtainsthe power spectrum of $\mathcal{R}$. For example, in the slow-roll approximation (when the scalar field energy is dominated by the potential),

$$P_{\mathcal{R}}(k) = \left(\frac{H^2}{2\pi (\bar{\varphi}')^2} \right) \Big|_{k=aH} \tag{1.8}$$

$$\approx A_s k^3 \left(\frac{k}{k_0} \right)^{n_s - 1}, \tag{1.9}$$

where the k-dependence of the first line is due to evaluation for a particular mode at the horizon scale - this introduces a scale dependence to the power spectrum. Here we arrive at an expression containing our last two parameters A_s and n_s, quantifying the amplitude of primordial fluctuations and the leading (in k) part of their scale dependence. The question now remains how to connect perturbations in $\mathcal{R}$ to those of the observed galaxy density at late times that traces large-scale structure. Neglecting the process by which fluctuations in $\mathcal{R}$ are converted into fluctuations in Standard Model species [6, 7, 8, 9], we now proceed to our next equation needed to construct the time evolution of the universe, the Einstein-Boltzmann equation for those species.

---

[6]Non-negligible higher-order correlation structure arises in specific inflationary models, and the phenomenological study of this behavior is termed Primordial non-Gaussianity, which we will see a bit of in Chapter 4.
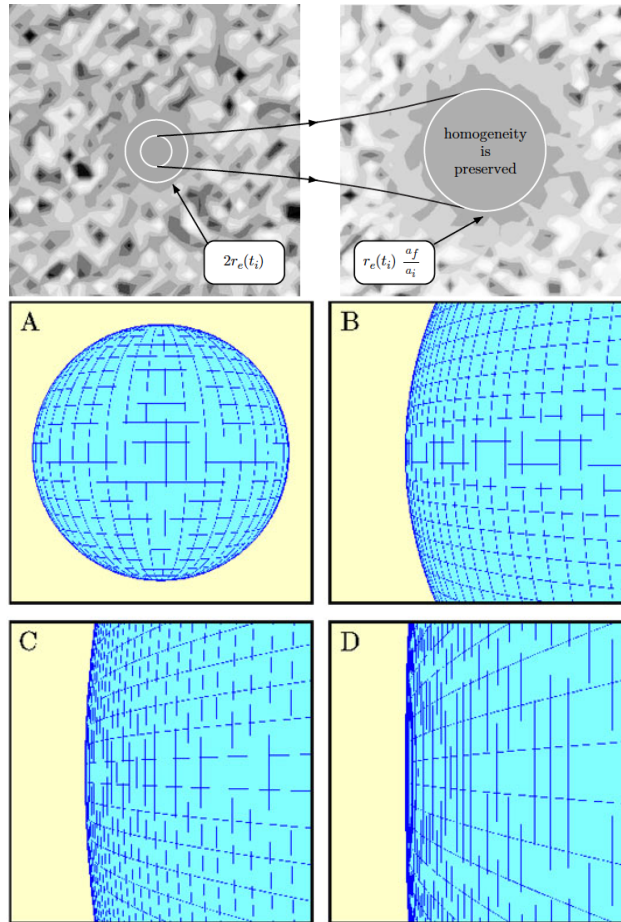
Figure 1.1: Illustration of how inflation solves the horizon (*top*) and flatness (*bottom*) problems. *Top*: A small initially homogeneous and isotropic region of size $r_e(t_i)$ is stretched by a factor $\frac{a_f}{a_i}$ during inflation, leading to homogeneity and isotropy on large-scales. *Bottom*: Schematically, a region of large initial spatial curvature (A) is stretched such that the region tends toward flatness (D). Taken from Refs. [5, 10].

1.2   Linear evolution

The primary quantities of interest in modern cosmology are statistics of the cosmic microwave background (CMB) and the matter power spectrum $P(k)$. These are sourced by fluctuations in the photon, CDM, baryon, and to a lesser extent, neutrino and potentially more exotic particle fields during matter domination - the CMB at a scale factor of $a \sim 10^{-3}$ and $P(k)$ at $a \lesssim 10^{-1}$. These sourcing fluctuations must evolve from the (assumed adiabatic) initial condition fluctuations generated by inflation that enter the horizon[7] during either

---

[7]For a detailed discussion of horizon crossing of individual Fourier modes of various fields, see Ch. 8 of Ref. [12].
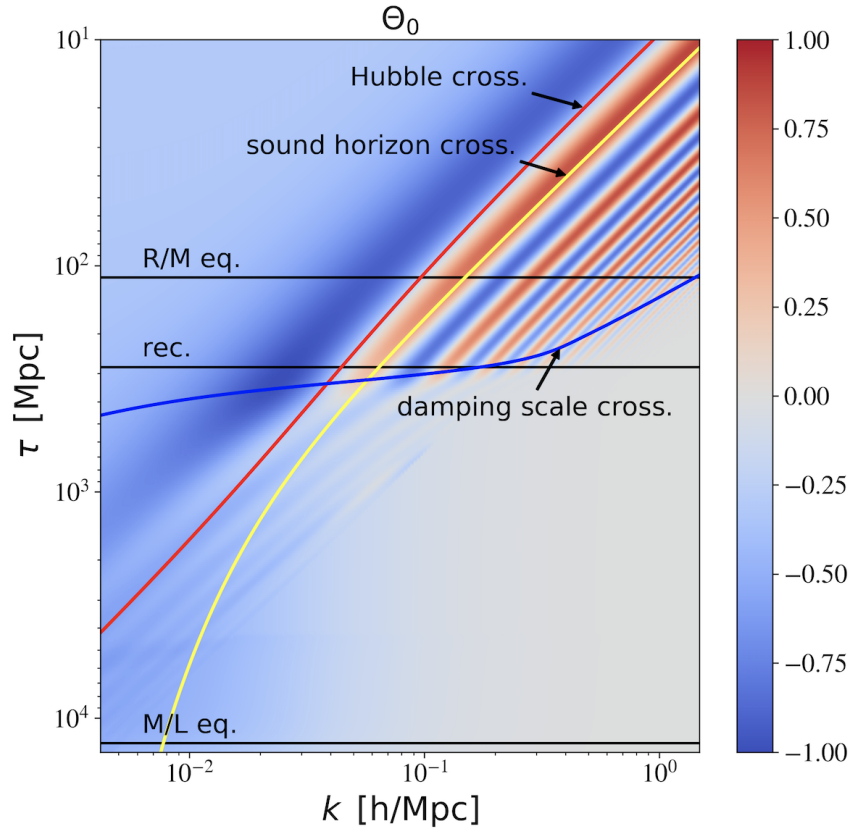
Figure 1.2: The linear evolution of the photon temperature monopole perturbation Θ_0 as a function of time (conformal time τ) and scale (wavenumber k). While especially apparent in the CMB, the memory of the photon-baryon oscillations is imprinted on the distribution of LSS at late times through the BAO feature. Various scales governing the generation and damping of fluctuations are highlighted. Produced using `CLASS` [11], adapted from here.

radiation (on smaller scales) or matter domination (on larger scales). Since we are not concerned with the evolution of arbitrarily small scale fluctuations, it suffices to consider the initial conditions of individual fields in radiation domination to solve for the evolution of the Einstein-Boltzmann system to obtain late-time fluctuations at all times relevant for observations.

Depending on the particle nature of each individual field, it will have a (assumed 1-particle) distribution function $f = f(\mathbf{x}, \mathbf{p}, t)$ on phase space typically described by Fermi-Dirac or Bose-Einstein statistics. The particle mass of each field and its possible interactions with other fields then, when combined with the Einstein equations, dictates the content of the Einstein-Boltzmann system of equations.

We will briefly state the general form of this system, but the full details of this Einstein-Boltzmann treatment for ΛCDM species (with the addition of massive neutrinos) can be

found in Ref. [13][8]. In the notation there, the Einstein-Boltzmann system can be written for an arbitrary species at linear order in terms of linear perturbations $\Psi = \Psi(\mathbf{x}, \mathbf{p}, t)$ to the background distribution function $f_0 = f - \Psi$ of the species for a given Fourier mode $\mathbf{k}$

$$\frac{\partial \Psi}{\partial \tau} + i\frac{q}{\epsilon}(\mathbf{k} \cdot \hat{n})\Psi + \frac{d \ln f_0}{d \ln q}\left[\phi' - i\frac{\epsilon}{q}\left(\hat{k} \cdot \hat{n}\right)\psi\right] = \frac{1}{f_0}\left(\frac{\partial f}{\partial \tau}\right)_C,$$
(1.10)

where, $q = ap$ is the comoving momentum magnitude (where p is momentum magnitude), ϵ is the comoving energy $\epsilon^2 = q^2 + m^2 a^2$ for a species with particle mass m, $\hat{n}$ is the momentum direction unit vector, and the right-hand side contains all possible collision terms (e.g. due to photon-baryon interaction), and we have written the equation in conformal Newton gauge[9], i.e., for a perturbed metric

$$ds^2 = a^2(\tau)\left[-(1 + 2\psi)\,d\tau^2 + (1 - 2\phi)\,dx^i dx^i\right].$$
(1.11)

The evolution of this system leads to a rich picture of the physics at play in the evolution of perturbations, and example of which is given for photons in Figure 1.2.

For large-scale structure, the case of the evolution of the "cold" species (CDM and baryons) during matter domination is of particular interest. For the moment we will neglect massive neutrinos, but see Chapter 3 for a more complete discussion that includes them. For these cold species, higher velocity moments are negligible, and the two species can each be modeled as a fluid. After photons decouple from the baryons around the time when the CMB we see is emitted, the cold species evolve together under the influence of gravity and any interactions between species can be neglected. At this time, the Einstein-Boltzmann system simplifies to the collisionless linearized Vlasov-Poisson equation by setting the right-hand side of eqn. 1.10 to zero. During this time of purely gravitational evolution, the baryons imprint the memory of the photon-baryon interaction at the time of recombination (the Baryon Acoustic Oscillations, or BAO[10]) on the total matter overdensity $\delta_m(k, a) = \frac{\Omega_c}{\Omega_b + \Omega_c}\delta_c(k, a) + \frac{\Omega_b}{\Omega_b + \Omega_c}\delta_b(k, a)$ where $\delta_X = \delta_X(\mathbf{x}, t) = \frac{\rho_X(\mathbf{x}, t)}{\bar{\rho}_X(t)} - 1$, which results in oscillations in wavenumber in the matter power spectrum.

After these simplifications, by taking velocity moments of the Vlasov-Poisson equation we obtain the familiar fluid equations augmented by the Poisson equation

$$\frac{\partial \delta}{\partial \tau} = -\theta$$
(1.12)

$$\frac{\partial \theta}{\partial \tau} = -aH\theta - \nabla^2\Phi,$$
(1.13)

---

[8]See also our open-source implementation of `Bolt`, the first differentiable linear Einstein-Boltzmann solver.

[9]For a nice introduction to various gauges, see here

[10]The BAO is a robust source of cosmological information - specifically in terms of the ΛCDM parameters Ω_c and h, and has a rich history in the development of LSS - see, e.g., here and here for quick physical overviews.

where $\theta = -\nabla \cdot \mathbf{v}$ and $\Phi = \phi$ in matter domination. For an individual Fourier mode, the dominant growing mode solution of the total matter overdensity $\delta_m(k, a)$ evolves in a scale-independent manner, and its time dependence is usually parameterized in terms of the "growth factor" $D(a)$, as $\delta_m(k, a) = D(a)\delta_m(k, 1)$. This linear description of the evolution of cold matter holds on large scales, and serves to set the initial conditions of nonlinear growth on smaller scales, which we address in the next section.

1.3 Nonlinear structure formation
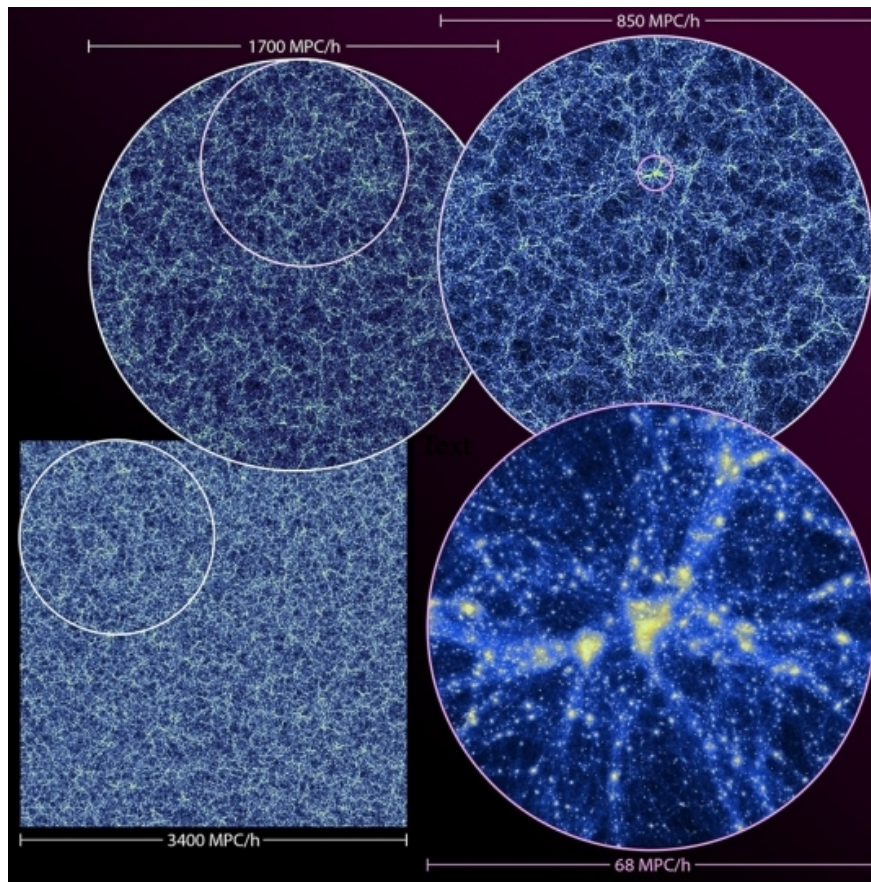
Figure 1.3: Several zoom-ins of a large-volume dark-matter-only cosmological N-body simulation. The illustration makes visually apparent the computational demand of simulating large volumes and high resolution, as well as the non-linear structure of the density field that is demonstrated through web-like structures. Taken from Ref. [14], where the simulation this visualization is based on is described further.

The distribution of matter that constitutes large-scale structure is not well-described by

linear perturbation theory except on the largest scales. The extremely low-density regions, as well as the very dense sheets, filaments, and highly-clustered nodes (see Figure 1.3) are all features that require solving a partial differential equation system to model - the fully nonlinear analog of the Einstein-Boltzmann equation system - the collisionless Vlasov-Poisson system:

$$\frac{\partial f}{\partial t} + \frac{p_i}{ma^2}\frac{\partial f}{\partial x^i} - m\frac{\partial \Phi}{\partial x^i}\frac{\partial f}{\partial p_i} = 0 \tag{1.14}$$

$$\nabla^2\Phi = \frac{3}{2}\frac{H_0^2}{a}\Omega_m\delta \tag{1.15}$$

$$\rho = m\int d^3 f(\mathbf{x}, \mathbf{p}, t), \tag{1.16}$$

where the comoving density is estimated using density tracer particles of mass m. Simplified models for this fully non-linear phase space evolution generally proceed by making several assumptions to reduce the complexity of the problem. Perturbative techniques solve the fluid equations that result from velocity moments of the Vlasov-Poisson equation [15].

By considering the matter density field as a cold fluid, we can replace the linear equation 1.12 with the fully nonlinear fluid equations by taking velocity moments of the Vlasov-Poisson equation:

$$\frac{\partial \delta}{\partial \tau} = \nabla\left[(1+\delta)v\right] \tag{1.17}$$

$$\frac{\partial v_i}{\partial \tau} = -aHv_i - v_j\nabla^j v_i - \nabla_i\Phi. \tag{1.18}$$

By expanding the density field as a power law in the linear solution as $\delta = \delta^{(1)} + \delta^{(2)} + ...$, and doing the same for the velocities (usually for the velocity divergence θ), a perturbative solution for these fields and their statistics (such as the power spectrum) can be found. For example, the matter density contrast in Eulerian perturbation theory at second order in δ is

$$\delta^{(2)}(\mathbf{x}) = \frac{17}{21}\left[\delta^{(1)}\right]^2 + \frac{2}{7}\left[s_{ij}^{(1)}\right]^2 - \psi_i^{(1)}\partial^i\delta^{(1)}(\mathbf{x}), \tag{1.19}$$

where s_{ij} is the tidal tensor with $s_{ij} = \left(\frac{\partial_i\partial_j}{\nabla^2} - \frac{1}{3}\delta_{ij}^{(K)}\right)\Phi$, and ψ is the linear displacement. Here a separability between the time and space dependence of the linear matter density contrast growth is assumed, with powers of the linear growth factor $D(a)$ carrying the time dependence (though this is broken in detail, e.g. by the presence of massive neutrinos).

Alternatively, one can take a Lagrangian approach to perturbative solution of the fluid equations - effectively following the ballistic trajectories of tracer particles. The leading order Lagrangian expression for the matter overdensity field is given by the Zel'dovich approximation [16, 17], in which the Lagrangian matter is displaced according to (in Fourier space)

$$\Psi^{(ZA)}(\mathbf{k}) = \frac{i\mathbf{k}}{k^2}\delta^{(1)}(\mathbf{k}), \tag{1.20}$$

and positions of initial tracer particles $\mathbf{q}$ are related to the late-time positions as $\mathbf{x} = \Psi^{(ZA)} + \mathbf{q}$. The Lagrangian picture can be extended to arbitrary order in displacement by using perturbative kernels [18]. Here we will not discuss Effective Field Theory (EFT) methods for perturbative modeling of LSS, but simply note that these methods are indispensable for modern perturbative analyses of LSS data, and direct the reader to Ref. [19] for an introduction.

Any perturbative solution is only valid when the expansion parameter is small, and, on small scales, fluctuations in the matter overdensity field are quite large. This breakdown of perturbation theory can be addressed through direct simulation of non-linear matter dynamics through numerical solution of the collisionless Vlasov-Poisson equation via dark-matter-only N-body simulations [20]. Cosmological simulations of this kind rely on Monte Carlo tracer particle draws of the density field, which are then used to estimate the gravitational potential. The potential is then used to estimate the gravitational force acting on the particles, which are moved to new positions, and the whole process is iterated until the simulation completes according to some stopping criterion. The starting point for such simulations is the previously mentioned Zel'dovich approximation (eqn. 1.20) applied to the linear density field (as computed by an Einstein-Boltzmann ODE system solver), or, more commonly, a higher-order extension of the ZA such as second order Lagrangian perturbation theory (2LPT).

Simulations that include gas dynamics and radiative transfer in addition to the dark matter dynamics are largely computationally infeasible for cosmological applications simply due to the dynamic range of length and time scales required for self-consistent numerical simulations thereof. While the future will invariably bring improvements in such simulations, a two-step process of simulating dark matter dynamics and applying corrections, e.g. due to baryonic/gas physics, seems likely to drive cosmological inference in the near future. We will see this issue again in Chapter 2.

The entirety of the above discussion has related only to an accurate description of the cold matter field. However, what survey instruments actually observe are photons emitted by tracers of large scale structure rather than the more easily described matter field. To learn about cosmology from observed photons, we need a model for connecting the underlying cold matter field to these photons.

1.4 Observed tracers

The nonlinear matter density field as modeled by N-body simulations cannot be observed with telescopes - instead they see galaxies that form in locally overdense regions (see Figure 1.4). A dark matter halo is roughly the overdense region immediately surrounding a galaxy (or another LSS tracer), and halos are therefore key for understanding the connection between the nonlinear density field and observed tracers. To understand where halos

---

[11]A more up-to-date map of observed galaxy positions can be found here or here.
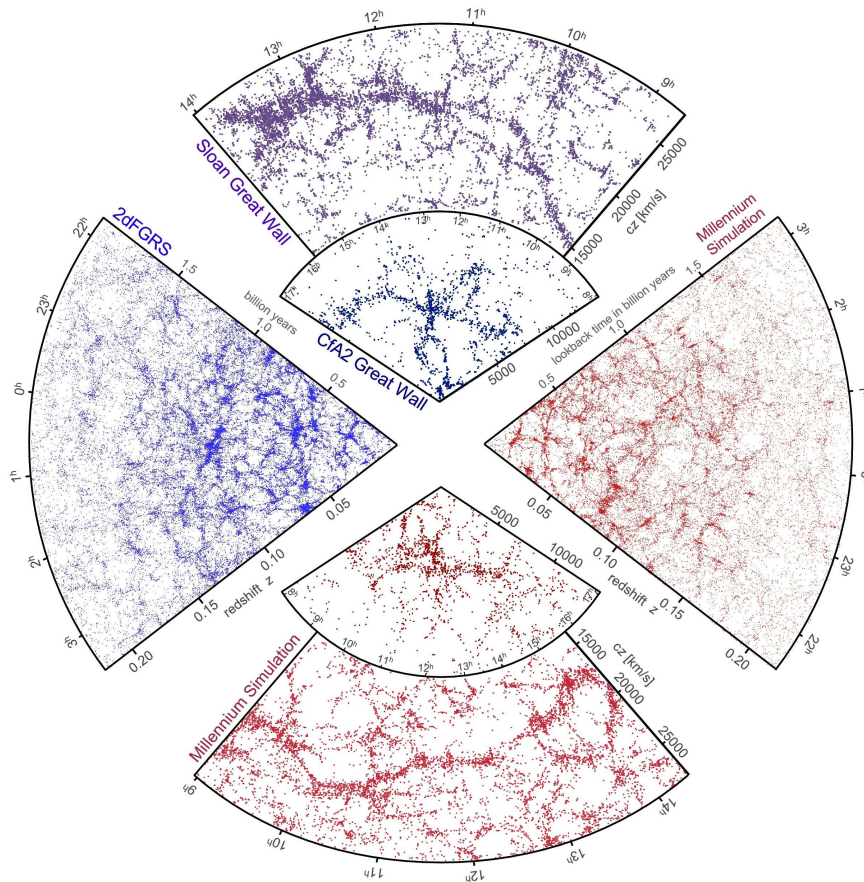
Figure 1.4: The distribution of observed and modeled large-scale structure tracers. Galaxies measured by several earlier redshift surveys (blue, purple) and mock galaxies populated into dark matter halos produced through simulations (red). Taken from Ref. [21, 22][11]

come from, it is instructive to consider the case of a single patch of cold matter collapsing under Newtonian gravity in an expanding FLRW background. Under the assumption of spherical collapse, when the growth of a matter perturbation within the patch exceeds a certain critical overdensity, the matter begins to collapse inwardly in a nonlinear fashion. Eventually, the collapsing matter virializes, leading to a dense halo whose size can be connected to its mass. This is an extremely simplified description of the process of halo formation, and we refer the reader to Refs. [23, 24] for more thorough discussion. This simple picture of a single halo's formation can be extended to a model for the statistical properties of the non-linear matter density field.

The halo model [25] assumes that all matter lives in dark matter halos to a decent approximation and constructs statistics of the matter density field out of several ingredients describing a population of halos. In the simplest case where only the mass (M) dependence

is modeled, these include the halo population number density (or abundance) $\bar{n}_h(M)$, the halo profile $\rho_p(r, M)$ as a function of radius r, and the halo bias $b(M)$. Then the matter power spectrum $P_{mm}^{(HM)}(k)$ is computed via integrals over the halo population

$$P_{mm}^{(HM)}(k) = P_{1h}(k) + P_{2h}(k) \tag{1.21}$$

where, roughly, the "2-halo" term describes the large-scale power of a density field consisting of a population of halos and the "1-halo" term captures the impact of the distribution of matter within halos. We discuss this model in more detail in Chapter 2.

The halo model connects the nonlinear matter density to collapsed structures that can host visible tracers, but does not describe the process of visible tracer formation itself. For the case where galaxies are the visible tracers of interest, this missing "galaxy-halo connection" must be modeled separately (see Ref. [26] for a review). This can be done at several levels, ranging from fully empirical statistical models to hydrodynamical simulations that attempt to numerically simulate the process of galaxy formation directly. One particularly general option is a symmetries-based bias expansion, or simply "bias models".

Modern tracer bias models provide a framework for precision cosmology that respects certain basic symmetries baked into the formalism (see Ref. [21] for a review). Following from the equivalence principle and Gaussian initial conditions, the large-scale bias expansion supposes that tracer formation is an approximately local process, and the number density of galaxies can depend on any scalars constructed from second derivatives and beyond of the (Newtonian) potential $\Phi(\mathbf{x})$, for example from its curvature $\delta \propto \nabla^2\Phi$ and the tidal tensor s_{ij}. In its general form, the tracer overdensity field δ_g at any order in the bias expansion can then be written as a sum over a finite number of (gravitational) operator fields $\mathcal{O}$ weighted by the scalar bias coefficients $b_{\mathcal{O}}$

$$\delta_g(\mathbf{x}) = \sum_{\mathcal{O}} b_{\mathcal{O}} \mathcal{O}(\mathbf{x}). \tag{1.22}$$

For example, up to quadratic order, we have the Eulerian bias expansion

$$\delta_g(\mathbf{x}) = b_1 \delta^{(2)}(\mathbf{x}) + \frac{b_2}{2} \left(\delta^{(1)}\right)^2(\mathbf{x}) + b_{s^2} s^2(\mathbf{x}) + (\text{h.d.}) + (\text{stoch.}), \tag{1.23}$$

where "h.d." refers to higher derivative terms related to the finite size of tracers, "stoch." refers to stochastic terms uncorrelated with the long-wavelength operators, and $\delta^{(2)}$ is the second order matter density contrast in Eulerian perturbation theory (eqn. 1.19). For further details see Ref. [21]. We will discuss bias models, the halo model, and the connection between observed tracers and dark matter halos in Section 2 and again in the context of primordial non-Gaussianity in Chapter 4.

To sum up, the essence of a physical model consistent with precision cosmological observations can be largely boiled down to the composition of:

1. the smooth background expansion history, expressed by eqn. 1.5

2. the evolution of curvature/inflaton fluctuations from Gaussian initial conditions (eqn. 1.7) during inflation to horizon crossing

3. the evolution of small (linear) fluctuations of SM species from the time of horizon crossing to the onset of gravitational nonlinearity (eqn. 1.10)

4. the evolution of larger fluctuations, either modeled perturbatively (eqn. 1.17) or via N-body simulations (eqn. 1.14)

5. a model for connecting the nonlinear matter distribution to observed tracers, such as galaxies (e.g., eqn. 1.22).

With a physical model in hand, it is then possible to obtain point estimates of cosmological parameters and uncertainty quantification of those estimates through statistical inference.

1.5 Inference
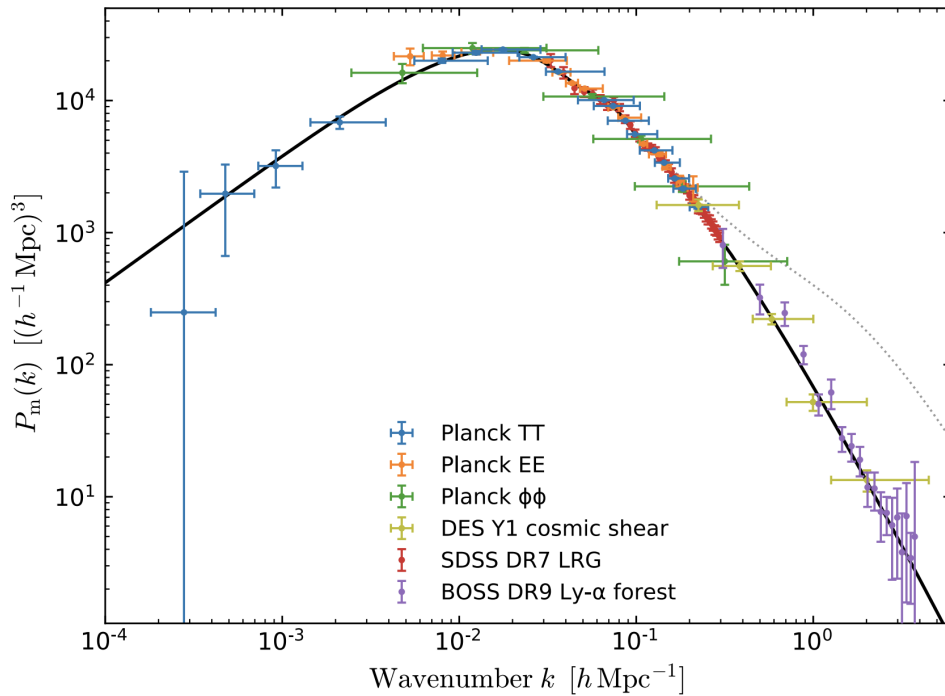
Figure 1.5: The linear matter power spectrum, as probed at a range of scales by the cosmic microwave background and several large-scale structure tracers - namely, CMB lensing, gravitational lensing of galaxy shapes, luminous red galaxy clustering, and Lyman-α forest flux correlations. Taken from Ref. [27][12].

The story of the universe told by our physical models is only credible due to the match of their predictions of key observables to data. Figure 1.5 shows one such observable, the matter power spectrum, as probed through the anisotropy power spectrum of the CMB as observed by Planck [28] and several LSS tracers of the matter density field. The excellent match between the points and the model curve reflects the status of the ΛCDM model as the touchstone of cosmology for the last decade. Observations of supernovae constitute another key pillar of modern cosmology (for some recent measurements see [29, 30, 31]) - the measurement of distances to supernovae was responsible for the discovery of dark energy (Λ), and has been corroborated by more recent measurements of baryonic acoustic oscillations in the matter power spectrum [32].

Constraints on model parameters (such as those of ΛCDM) are obtained through comparing the data to models via Bayesian inference [33]. As a brief overview, given a parametric model $m(\theta)$ and data d, the goal of inference is to obtain the (usually marginal) probability distribution of θ given the data d - i.e. the form of $p(\theta|d)$, the posterior. Bayes theorem gives a way to obtain this posterior

$$p(\theta|d) = \frac{p(d|\theta)}{p(d)}p(\theta),\tag{1.24}$$

in terms of the functions $p(d|\theta)$ (likelihood), $p(\theta)$ (prior), and $p(d)$ (evidence). Ordinarily, we are not interested in the normalization of the posterior $p(d)$, and need only to find a way to supply $p(d|\theta)$ and $p(\theta)$. Most commonly, $p(d|\theta)$ is assumed to have a known (usually Gaussian) form $p(d|\theta) = L(d, \mu, C|\theta)$ where the mean μ is usually zero, and the covariance C is either known analytically, generated from many realizations of synthetic data, or estimated from the data itself, i.e. via jackknife resampling. The prior, $p(\theta)$, must be supplied by the person doing the inference based on domain knowledge.

The process of inference proceeds generally in two steps - the first is a determination of the model parameters θ that best characterize the data (according to either the likelihood or posterior) and the second is characterizing the uncertainty on the estimate of θ. The first step is achieved through optimization algorithms, which use the value of the function in question (e.g. likelihood or posterior) as well as potentially other information (like the gradients of that function) to find the optimal value of θ through a series of steps. With this optimum in hand, it can be used as a starting point for quantifying the uncertainty on θ.

In practice, for any reasonable number of parameters θ, obtaining parameter uncertainty estimates by working directly with high-dimensional distributions like $p(\theta|d)$ is classically intractable. Instead, we can obtain an approximation of the posterior through Markov Chain Monte Carlo (MCMC) sampling, a recipe by which parameters θ are drawn starting from the prior and weighted by the likelihood through a procedure that eventually produces posterior samples - the details of which have been the subject of a vast literature going back 70 years [34]. The so-obtained posterior samples can then be used to estimate the

---

[12]This image is already quite out of date - see Ch. 25 of Ref. [1] for a compilation of somewhat more recent LSS measurements in the less pedagogical context of cosmological parameter measurements.

(usually unnormalized) density of the posterior, its marginal distributions, or other statistics of interest. Increasingly in cosmology, alternatives to this process of sampling are being considered in the literature due to advances in computation and machine learning [35, 36], but sampling methods remain central to most cosmological data analyses. We will discuss a procedure for optimization that makes use of a machine learning method to approximate a target distribution in Chapter 5.

1.6   Outline

The history of the universe that the community has constructed is not only gratifying to behold, but also physically diverse - relying on the language of general relativity, quantum field theory, and statistical and fluid mechanics. Numerical methods and phenomenological models are required to carry this foundational language up to the high precision of modern observations of the sky. This thesis touches on several developments of said tools, specifically, with a focus on using modern statistical methods to sharpen existing models. These additions are especially relevant when dealing with physical effects that are relatively uncertain - primarily as related to the galaxy-halo connection and the general process of non-linear structure formation.

In detail, Chapter 2 develops a phenomenological model of tracer and matter power spectra aimed at approximately capturing both of these complications, while Chapter 3 relates careful numerical treatment of nonlinear structure formation with N-body simulations in the presence of massive neutrinos. The uncertainty of the galaxy-halo connection also motivates empirical models for connecting galaxy populations in surveys with their underlying dark matter halos, for which cosmological information is more accessible. Chapter 4 outlines a method for using a machine learning method trained with hydrodynamical simulation information as such an empirical model in the context of primordial non-Gaussianity.

Finally, the precision of observations leads to a large number of input parameters needed in suitable models. This begets a computational problem, as forward models even for the usual compressed statistics (such as the power spectrum of galaxies) require of order several seconds to evaluate, and must often be evaluated hundreds of thousands or millions of times for parameter inference. One strategy for addressing this problem is to use approximate models at the the level of the probability distribution of interest (the likelihood $p(d|\theta)$ or the posterior $p(\theta|d)$), and optimize these to jump-start the inference process before performing costly corrections using the exact distribution. Such a strategy is the subject of Chapter 5. The application of these methods makes a small but significant stride toward preparing for a landscape of survey data that is richer and vaster than ever before.

Chapter 2

Halo Zel'dovich Perturbation Theory

In this chapter, we update Halo Zeldovich Perturbation Theory (HZPT, [37]), an analytic model for the two-point statistics of dark matter, to describe halo and galaxy clustering, and galaxy-matter cross-correlation on nonlinear scales. The model correcting Zeldovich has an analytic Fourier transform, and therefore is valid in both configuration space and Fourier space. The model is accurate at the 2%-level or less for P_{mm} ($k < 1$ h/Mpc), P_{hm} ($k < 1$ h/Mpc), P_{hh} ($k < 2$ h/Mpc), P_{gm} ($k < 1$ h/Mpc), P_{gg} ($k < 1$ h/Mpc), ξ_{mm} ($r > 1$ Mpc/h), ξ_{hm} ($r > 2$ Mpc/h), ξ_{hh} ($r > 2$ Mpc/h), ξ_{gm} ($r > 1$ Mpc/h), ξ_{gg} ($r > 2$ Mpc/h), for LRG-like mock galaxies. We show that the HZPT model for matter correlators can account for the effects of a wide range of baryonic feedback models and provide two extended dark matter models which are of 1% (3%) accuracy for $k < 10$ (8) h/Mpc. We explicitly model the non-perturbative features of halo exclusion for the halo-halo and galaxy-galaxy correlators, as well as the presence of satellites for galaxy-matter and galaxy-galaxy correlation functions. We perform density estimation using N-body simulations and a wide range of HOD galaxy mocks to obtain correlations of model parameters with the cosmological parameters Ω_m and σ_8. HZPT can provide a fast, interpretable, and analytic model for combined-probe analyses of redshift surveys using scales well into the non-linear regime.

2.1   Introduction

The goal of large-scale structure (LSS) analysis is to extract cosmological information from the nonlinear matter density field. Nearly all modern cosmological analyses are built upon two-point statistics that probe this field [38, 39]. Increasingly precise measurements of two-point correlators in modern galaxy surveys demand percent-level accuracy of theoretical models of these correlators. This is especially true on smaller scales where the effect of

survey sample variance is minimal and non-perturbative effects dominate. However, models of two-point statistics require a tradeoff between the scales they access and the amount of theoretical control they have.

Perturbation theory (PT) provides an analytic model of the density field on large scales [15]. Two-point correlators in perturbative models are limited by the scale at which nonlinear effects dominate the dark matter dynamics which is a much larger scale than the minimum scale to which current surveys are sensitive [40]. However, PT models remain attractive due to the control over theoretical errors they afford in their domain of validity. Extensions of perturbation theory based on an effective fluid description of the density field (EFT) have pushed deeper into the quasi-linear regime [41, 42, 43, 44, 45]. Such extended perturbation theory models have recently been used successfully for analysis of cosmological parameters, though with nuisance parameters that are fitted to numerical simulations (e.g. [46, 47, 48]). However, there is a limit to any perturbative model, even in the EFT framework, as non-perturbative effects and nonlinear gravitational evolution dictate the behavior of the density field on scales less than a few Mpc/h. In fact, it is clear that perturbation theory does not actually converge to the fully nonlinear result on smaller scales, at least in one dimension, at infinite order [49, 50]. This is due to the fundamentally non-perturbative nature of small-scale dark matter dynamics.

An alternative analytic model that includes non-perturbative effects in the form of halo formation is the halo model [51, 52, 25, 53]. The halo model assumes that all dark matter is tied up in gravitationally bound, non-overlapping halos, which have a prescribed density profile and an abundance set by the halo mass function. Two-point matter correlators are computed by way of mean halo profiles integrated over the halo mass function and halo bias. The halo model has seen success in the last few decades, and is used in modern analyses, albeit usually with some modifications, to model fully nonlinear scales (e.g. [54]). However, the halo model struggles in the so-called "transition regime" between the one-halo and two-halo terms. The halo model also fails to ensure large-scale conservation laws are satisfied, and as a result cannot be completely correct in its original form [55]. Despite the successes of both PT and the halo model in complementary regimes, it is clear that neither of these analytic models alone are adequate to fully describe the nonlinear density field.

Without sufficiently accurate analytic models of matter clustering, simulations can instead act as a model of nonlinear dynamics. N-body simulations provide Monte Carlo realizations of Newtonian dynamics in the fully-nonlinear regime, producing a nonlinear matter density field for a given cosmological model. As with PT and the halo model, N-body simulations are limited by a fundamental assumption - namely that evolution of the matter distribution is fully described by collisionless cold dark matter obeying Newtonian gravity. However, this assumption does not limit the scales accessible to the model or the types of nonlinear structures it can produce, which are constrained only by numerical resolution. Recent advances in computing have led to the rise of large-volume, high-resolution simulations (e.g. [56, 57, 58, 59, 60]), albeit with questions of convergence at the percent-level [61]. With these have come approximate methods of simulation that aim to obtain comparable solutions with much less computation time [62, 63, 64]. As redshift surveys push to larger volumes

and higher number density, the intractability of running many sufficiently resolved simulations at multi-Gpc volumes (necessary for capturing large modes and estimating covariance) to produce two-point statistics has motivated various fitting functions and interpolations of two-point statistics produced by high resolution simulations (e.g. [54, 65, 66, 67, 68]). At the smallest scales probed by observations, baryonic effects on the matter distribution are also a concern [69, 70], and to properly simulate their impact on large-scale structure requires a full understanding of galaxy formation and hydrodynamic simulations that include feedback [71, 72]. There have been some recent efforts to correct for baryons and mitigate this issue by modifying the output of dark-matter-only simulations [73, 74, 75].

In practice we do not observe the nonlinear matter density field, but instead its tracers. Modeling the connection between tracers and the underlying density field is a complex task, and there are several prevailing approaches to this problem. The large-scale bias approach extends the philosophy of perturbation theory to parameterize the tracer field as a linear combination of locally-leading gravitational observables [40]. These models have been successful on large scales, but face the same issue of PT for modeling dark matter, namely the presence of a nonlinear scale that characterizes the dark matter dynamics [76]. On top of this fundamental limit, there is an additional limiting scale, the nonlocality scale, which characterizes the formation of the tracer (e.g. halos/galaxies) and is not necessarily coincident with the nonlinear scale. Recently, some pragmatic semi-analytic models have partially circumvented this issue by considering bias with respect to a fitting function for the non-linear matter power [77], using bias templates measured from simulations [78], or by extending the halo model via functions fit to simulations [79]. In the halo occupation model approach, luminous tracers are assigned to a catalog of dark matter halos via a prescription for stochastically populating the halos - this is the Halo Occupation Distribution (HOD) framework [80, 51, 81]. HOD models, usually combined with halos found in N-body simulations, are used for some modern analyses that include small-scale galaxy clustering [54, 82]. There are also more involved models of the galaxy-halo connection that are less frequently used in cosmological analyses [26]. Finally, hydrodynamic simulations that include baryonic/gas physics attempt to model galaxy formation more directly, albeit with stochastic subgrid models mixed in [71]. These simulations are extremely computationally expensive to run, and typically cannot be run at the sufficient number of realizations or volume to be relevant for modern redshift surveys.

Given the limitations of existing models for the nonlinear density field and its tracers, a lofty goal is then to produce an interpretable, analytic model that is accurate across all scales of cosmological interest. Toward this goal we build upon the hybrid PT-halo model approach put forward by [83] (MS14) and [37] (SV15) for modeling dark matter two-point correlators - Halo-Zeldovich Perturbation Theory (HZPT). There have been several other efforts in this hybrid-modeling direction for dark matter only correlators [84, 55, 85, 86]. However, both HZPT in its original incarnation and these previous works do not account for baryonic effects on the matter two-point correlators or the more observationally-relevant modeling of general tracer two-point correlators. We address these shortcomings in this paper.

The purpose of this paper is twofold. First, we aim to provide a fast, analytic, and

accurate model for two-point correlators of matter (accounting for baryonic effects) and tracers on small scales, which necessitates modeling of halo exclusion and satellite galaxies. In addition, we provide power-law fits for matter correlators, and a joint distribution of cosmology and HZPT parameters for describing LRG-type (mock) galaxies as the foundation for an emulator-like approach to analysis of two-point statistics. We restrict our attention to real space correlators, as the most immediate application of HZPT is to projected statistics in a "3x2pt"-style analysis.

We first describe the N-body simulations and HOD mocks used in this paper in Section 2.2. We review the HZPT model and outline its basic structures in Section 2.3. We apply the model to dark matter correlators and discuss the impact of baryons in Section 2.4. We apply the model to halos in Section 2.5, and provide a detailed discussion of halo exclusion before moving on to mock galaxies in Section 2.6 and concluding in Section 2.7.

2.2   N-body Simulations and Halo Occupation Distribution

CrowCanyon simulations and CM HOD mocks

We use particle output and halo catalogs from the `CrowCanyon` N-body simulations to test the HZPT model on matter and halo statistics. These simulations were run using FastPM [62] with $N_p = 6144^3$ particles, a box size of $L_{\rm box} = 3200$ Mpc/h, a boost factor $B = 2$ (for a 12288^3 PM force grid and Nyquist wavenumber $k_{\rm Nyq} \approx 12$ h/Mpc), using the Planck15 cosmological parameters [28] (without neutrino effects). `CrowCanyon` halos were identified using the `nbodykit` FoF halo finder [87] with linking length $b = 0.2$. We computed simulation power spectra using `nbodykit` [88] with a FFT mesh using $N_{\rm mesh} = 2048$ ($k_{\rm Nyq} \approx 2$ h/Mpc) using a correction for compensation [89], and a Triangular Shaped Cloud interpolation window. Similarly, we computed correlation functions using FFTs on large scales with $N_{\rm mesh} = 1024$, matched at $r = 10$ Mpc/h to the result of the `corrfunc` [90] pair counting algorithm (as included in `nbodykit`) with 100 logarithmically-spaced bins and a maximum bin of 10 Mpc/h. Power spectra are sample variance cancelled using unitary amplitude ("paired-fixed") power spectra [91] at the same random seed as the N-body simulation initial conditions. To compute the linear theory power spectrum we use `CLASS` [11]. To quickly compute the Zeldovich power spectrum we modify a version of the FFTLog-based code employed in [92] and [93]. Fits are performed using the `scipy` implementation of the "Trust Region Reflective" optimization algorithm.

We used the `nbodykit` [88, 94] implementation of the simple 5-parameter Zheng '07 HOD model [95] to populate `CrowCanyon` halos with galaxies. This implementation modulates the satellite occupation by that of the centrals, assumes no halo-central mis-centering, and places satellite galaxies in halos according to an NFW profile. We use 100 sets of HOD parameters sampled from a symmetric latin hypercube with a number density fixed to near the BOSS CMASS [96] value ($\bar{n}_g = 4.2 \times 10^{-4}$ h^3 Mpc^{-3}) at $z = 0.55$ (Fig. 2.1). The parameter

$\log M_{\min}$ is not drawn from the hypercube, and is instead fixed by integrating over the `CrowCanyon` halo mass function to reproduce the appropriate $\bar{n}_g$. The ranges of parameters considered are: $\alpha \in [0.5, 1.0]$, $\log M_1 \in [13.5, 14.5]$, $\log M_0 \in [11, 13.5]$, $\sigma_{\log M} \in [0.01, 0.8]$. This results in a large range of satellite fractions ($f_{\text{sat}} \approx 0.01 - 0.6$), which is discussed further in Section 2.6. We refer to this HOD mock galaxy sample as the "CM" sample since it approximates the BOSS CMASS galaxy number density, redshift, and roughly follows the HOD parameterization of the CMASS analysis of [82].
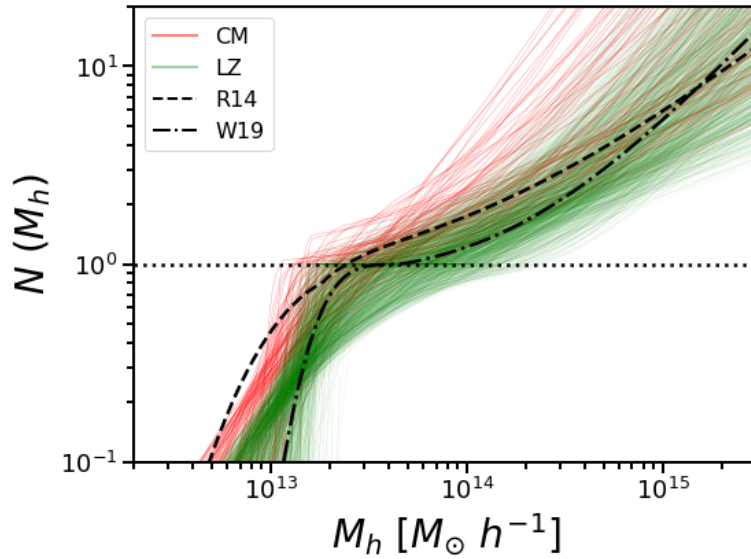
Figure 2.1: Halo occupation space. CM refers to the wider HOD sample space corresponding to the larger-volume `CrowCanyon` simulations, while LZ refers to the LOWZ-like HOD sample space corresponding to the Aemulus simulations. R14 and W19 refer to the halo occupations using the (fiducial) mean constrained HOD parameters (using the model of [95]) for the BOSS CMASS and LOWZ data from [82] and [54], respectively.

Aemulus simulations and LZ HOD mocks

We similarly generate LOWZ-like HOD mock catalogs from the Aemulus simulations [60]. We use 10 different cosmologies (Boxes 0-9) with 20 HODs per cosmology. We use a number density close to the LOWZ value ($\bar{n}_g = 3 \times 10^{-4} \ h^3 \ \text{Mpc}^{-3}$), and use a snapshot at $z = 0.25$. Halos are populated according to the 200b mass definition of the NFW radius and concentration. We computed simulation power spectra using `nbodykit` [88] with a FFT mesh using $N_{\text{mesh}} = 1024$ using a correction for compensation [89], a Triangular Shaped Cloud window, and interlaced anti-aliasing [97]. We extend the maximum `corrfunc` pair count bin to 20 Mpc/h and use an `FFTCorr` grid of 512^3 for correlation function measurements. We use a similar range of parameters as [54]: $\alpha \in [0.5, 1.5]$, $\frac{M_1}{M_{\min}} \in [2.5, 20.0]$, $\frac{M_0}{M_1} \in [0.0, 0.4]$,

$\sigma_{\log M} \in [0.01, 0.8]$. For each set of HOD parameters, the minimum halo mass parameter $\log M_{\min}$ is fixed to match the LOWZ number density once all of the other parameters have been chosen. This value is determined by an integral over the halo mass function - to account for variation of the mass function with cosmology, we use the Aemulus emulator for the halo mass function trained on the simulations [98]. We refer to this HOD mock galaxy sample as the "LZ" sample since it approximates the BOSS LOWZ galaxy number density, redshift, and HOD parameterization of the LOWZ analysis of [54]. Figure 2.1 shows the halo occupations corresponding to these parameters (as well as the same for the CM sample).

2.3 Review of Halo-Zeldovich Perturbation Theory

In this section, we review the HZPT model for two-point statistics as developed in MS14, SV15, and [99] (H17), briefly recounting the relevant aspects of the halo model. We discuss each of the terms, their PT/halo model origin, and the scales at which they are relevant. Previous iterations of the HZPT model were applied only to matter and halo-matter correlators, and discussion in this section is restricted to these models, though we expand upon them more generally for tracers in Sections 2.5 and 2.6.

The halo model

The halo model [51, 52, 25, 53] makes the assumption that all matter resides in virialized halos of mass $M_{\mathrm{vir}} = \frac{4}{3}\pi R_{\mathrm{vir}}^3 \Delta_{\mathrm{vir}} \bar{\rho}_m$, and splits the two-point statistics of the matter field into correlations between halos (two-halo term) and correlations within a single halo (one-halo term):

$$P_{mm}(k) = P_{1h}(k) + P_{2h}(k), \ \xi_{mm}(r) = \xi_{1h}(r) + \xi_{2h}(r). \tag{2.1}$$

The ingredients of the halo model are: the halo mass function $dn(M)$ with $n(M)$ the number density of halos at fixed mass M, along with halo bias $b(M)$, and the spherically averaged halo profile $\rho_M(r)$(e.g. NFW [100]). The usual halo-model one-halo and two-halo expressions are then:

$$P_{1h}(k) = \int dn(M) \frac{M}{\bar{\rho}} |u_M(k)|^2, \tag{2.2}$$

$$P_{2h}(k) = \left(\int dn(M) b(M) u_M(k) \right)^2 P_L(k), \tag{2.3}$$

with linear power $P_L(k)$, and where the configuration space quantities are given by the Fourier transform of eqns. 2.2 and 2.3. Here $u_M(k)$ is the Fourier transform of the density profile normalized by the mass enclosed in the halo:

$$u_M(k) = \frac{4\pi}{M} \int_0^{R_{\mathrm{vir}}} dr \ r^2 \rho_M(r) j_0(kr), \tag{2.4}$$

with j_0 denoting the 0th-order spherical Bessel function. The profile is usually parameterized in terms of a characteristic scale radius r_s, (which can be written in terms of the halo concentration $c_{\rm def}(M) = \frac{R_{\rm def}}{r_s}$):

$$\rho_M(r) = \frac{\rho_0}{\left(1 + \frac{r}{r_s}\right)^2 \left(\frac{r}{r_s}\right)}. \tag{2.5}$$

We only explicitly compute halo model quantities in Appendix 2.8, and in that case use the NFW concentration-mass relation of Ref. [101] and the 200c mass definition.

Two-halo - Zel'dovich

The HZPT model replaces the traditional halo model two-halo term (eqn. 2.3) with the Zel'dovich Approximation (ZA), or the leading-order Lagrangian perturbation theory (LPT) power spectrum[1] [16, 17, 102, 103]. The ZA provides a beyond-linear-theory description of large scales, including large-scale nonlinear bulk flows. A benefit of the ZA is that the IR resummation that would be necessary in Eulerian perturbation theory (SPT) or in an EFT extension thereof is not required at several-percent accuracy, [85, 76, 42], as the Baryonic Acoustic Oscillation (BAO) wiggles are already captured quite well by the ZA [45]. More importantly for the purpose of this paper, ZA provides a useful ansatz for halo model extension, since a compensated halo profile added to ZA provides a description which is consistent with perturbation theory in the regime of its validity on large scales (Section II of SV15). To address the known deficit in power in the ZA power spectrum on large scales, SV15 matched the ZA power spectrum to that of SPT to write the amplitude of the one-halo term that also describes small scales, which we now address.

One-halo - Broadband Beyond Zel'dovich

The central feature of the small-scale HZPT model is the Broadband Beyond Zel'dovich (BB) term. In its initial formulation (MS14), this term replaces the one-halo term (eqn 2.2) to express a contribution to the power that is provided by an expansion of the Fourier transform of the halo profile in even powers of comoving wavenumber k. The coefficients of this expansion can in principle be obtained by integrating r^{2n}-moments of a prescribed halo profile up to a chosen "halo radius" at which the profile is truncated, though this was not done in MS14 or SV15 and these coefficients were simply fitted to simulations (expressions given in Appendix 2.8). To prevent large-k divergences, SV15 took this expansion in even powers of k and replaced it with a Padé-type term:

$$P_{BB}(k) = A_0 F(k) \frac{\sum_{m=0}^{m=n_{\max}-1}(kR_m)^{2m}}{\sum_{n=0}^{n=n_{\max}}(kR_{nh})^{2n}} \tag{2.6}$$

---

[1]Using an alternative perturbative substitute for the two-halo term can also prove fruitful, and we provide an example in Section 2.4, but by default in this paper we stick with the ZA.

The Padé approximation to the the k^2 expansion improves the range of validity of the model greatly by forcing the expansion to smoothly transition to zero as $k \to \infty$. Not only does this resummation remove high-k divergence due to polynomial terms, but it also increases the maximum wavenumber k up to which the BB term is a decent approximation to the Fourier transform of an idealized halo profile, which also transitions to zero for large k. We note that in the EFT sense, the BB term is not "stochastic", since we do not enforce that it is uncorrelated with the ZA term. The two-point correlation function (2PCF) is given by the Fourier transform of this expansion, which is analytic[2], and an expression for which (for $n_{\max} = 0, 1, 2$) can be found in Appendix 2.8.

Compensation

A well-known limitation of the original halo model is that mass and momentum are not conserved on large scales [55]. In the $k \to 0$ limit, conservation of these quantities requires $\lim_{k \to 0} P_{mm} \propto k^4$ [104]. The violation of this requirement for the halo model arises from the k^0 contribution of the one-halo term, which is due to the Poisson contribution from a finite number of halos [51]. The HZPT model addresses this by multiplying the BB Padé expansion term by a compensation kernel $F(k)$, which suppresses the low-k k^0 contribution (though at leading order this term goes like k^2). It is possible to explore more complicated forms of the compensation kernel (MS14), or to compensate the functional form of the halo profile iteslf [105], but here we keep with the previous iterations of HZPT and use the simple Lorentzian kernel with a single parameter R:

$$F(k) = \left(1 - \frac{1}{1 + k^2 R^2} \right). \tag{2.7}$$

SV15 matched to SPT to find a value of the compensation parameter R that was in good agreement with simulation measurements ($R \approx 26$ [**Mpc/h**] at $z = 0$). We discuss compensation for tracer-matter cross-correlations in Appendix 2.9.

The full model

The contributions to the model from the ZA and different BB terms are shown in Figure 2.2. The parameters are: A_0, which is related to the one-halo amplitude $\bar{\rho}_m^{-2} \int dn(M) M^2$ and does not depend on the profile, the R_{nh}, which are associated to the r^{2n}-moments of the halo profile, and the compensation scale R. A detailed discussion of these terms is provided in Appendix 2.8. There we also provide a full review of the original machinery of MS14 for the expansion in even powers of k. In the main text we will take a pragmatic approach, always fitting for the model parameters.

---

[2]To clarify, we take "analytic" in the sense of closed-form (as in [51]), rather than in the technical sense of functions.
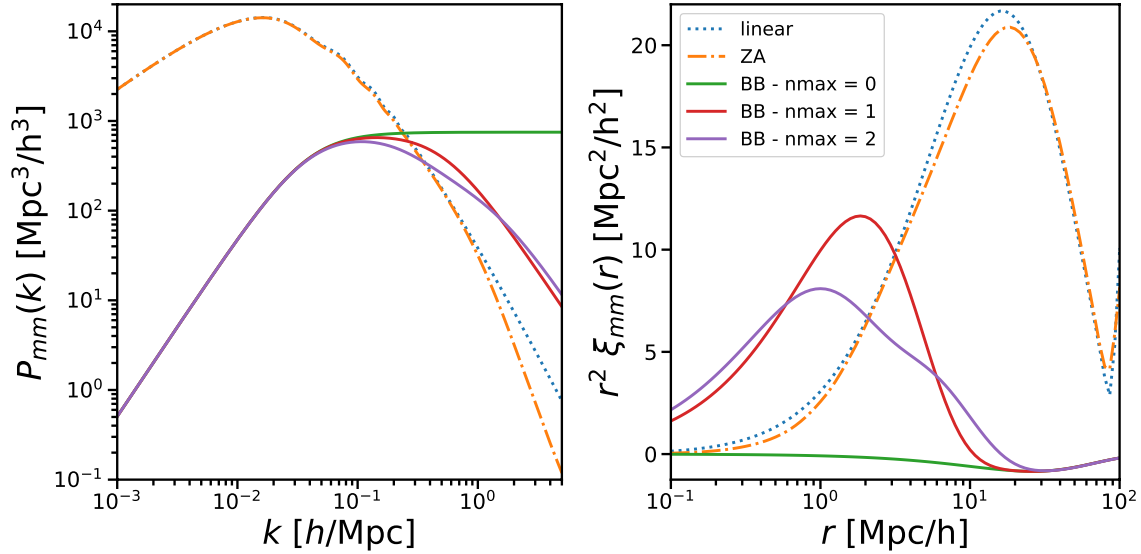
Figure 2.2: *Left:* Illustration of separate components of our model in Fourier space, as well as a comparison to linear theory. *Right:* The same components of the model in configuration space.

2.4 Matter Correlators & Baryons

We present the HZPT model for real-space two-point correlators of matter, including the effects of baryonic physics, which are most relevant for weak lensing observations. We review a calculation of MS14 of the profile expansion coefficients in the presence of a single model of AGN feedback, before performing an expanded calculation in the context of HZPT using several baryonic physics models. We also present a higher-n_{max} model that fits down to $k \approx 8$ $h/$Mpc at the 3% level, as well as a model with an augmented two-halo term that fits close down to $k \approx 10$ $h/$Mpc at the 1% level.

Before turning to baryons we briefly remark on the use of HZPT as a model of dark matter two-point correlators. Slightly generalizing the results of SV15, we show the fits over a range of redshifts ($0 \leq z \leq 2.5$) of the HZPT model to the power spectrum and correlation function in the `CrowCanyon` simulations in Figure 2.3. The fits for $P_{mm}(k)$ ($\xi_{mm}(r)$) are accurate to $\approx 2\%$ for $k > 1$ $h/$Mpc ($r < 1$ Mpc/h) for all but the highest redshift considered. Fits are performed in Fourier space, and these best-fit parameters are provided as input to the Fourier space model.

Modeling baryonic physics

To avoid biasing inferred cosmological parameters obtained through the matter two-point correlators, we must account for the effect of baryons on the matter distribution. Baryons
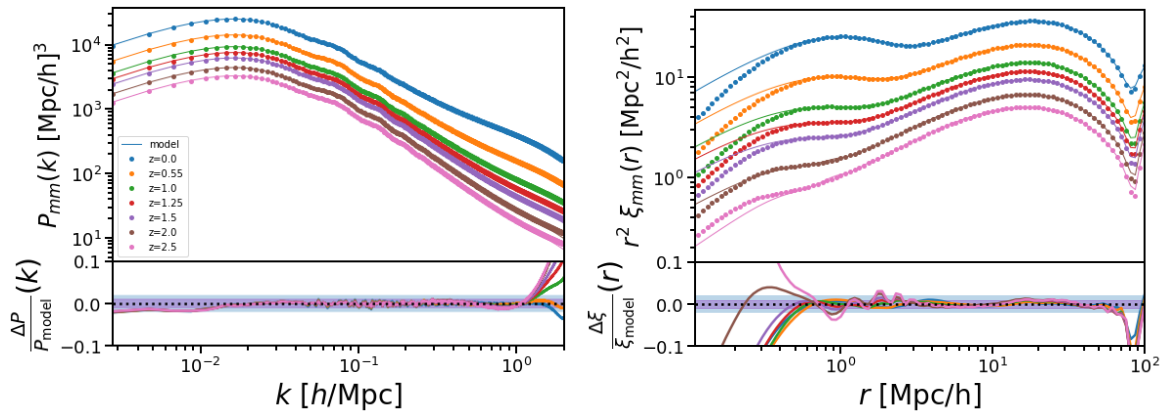
Figure 2.3: *Left:* Fits to the `CrowCanyon` matter power spectrum using the base HZPT model for a range of redshifts ($z = 0 - 2.5$). *Right:* The same for the matter correlation function. Colored bands are at 1% and 2%.

modify the dark-matter-only (DMO) halo profile in several ways, which has been explored in detail (e.g. [61, 71, 106, 75, 72]). In the presence of baryons, gas and stars must be accounted for in the halo profile in addition to dark matter. AGN and supernova feedback effects also move gas away from the halo center, which redistributes dark matter within the halo profile. From the perspective of the HZPT model, the effect of baryons should only be to modify the BB terms (see Appendix 2.8 for an illustrative calculation using halo profiles). Baryons should not affect the large scales relevant for the halo compensation R or the one-halo amplitude A_0, assuming conservation of mass between the DMO and DM+baryon scenarios (up to two-fluid corrections in the ZA [107, 108, 109, 110]). This is the same rationale used to motivate scale cuts (e.g. [111]). MS14 fitted changes in the power spectrum due to a single AGN model of feedback in the coefficients for the first three terms of the k^2 profile expansion up to $k = 0.8$ $h/$Mpc. They found that the change in A_0 (which is the same as our A_0 up to small changes due to the compensation term) is almost an order of magnitude lower than changes in the higher-order parameters, which change at the $\sim 5\%$ level, and if they fix A_0 the change in the other terms is larger but still effectively captured by the profile expansion.

We achieve a similar but improved result compared to MS14 using a more involved comparison. We use a larger k-range, fitting out to $k = 1$ $h/$Mpc using the matter power spectrum from simulations. We also employ a more diverse range of baryonic physics models by multiplying the dark-matter only `CrowCanyon` matter power at $z = 0$ by the ratio $P_{\mathrm{baryon}}/P_{\mathrm{DMO}}$ for 13 different models from [112, 113, 114, 115, 116, 117, 106] as used in [118]. Figure 2.4 shows fits to dark-matter-only power spectra and power spectra including the effects of baryons for several models with A_0 and R fixed to their DMO values. Clearly the $R_{n,nh}$ parameters are flexible enough to accurately account for feedback, and marginalizing over them should remove biases in cosmological parameter constraints. We note that while

all models are fit to the 1% level down to $k = 1$ $h/$Mpc, the simulations with the largest deviations in the R_{nh} parameters from the DMO case are Illustris and Horizon-AGN. This is the case for Illustris since it has been shown to have an unrealistically strong feedback model (compared to other hydrodynamic simulations) in terms of its effects on the power spectrum for $k \leq 1$ due to low baryon fraction (compared with the observed value in galaxy groups) in high mass halos [72]. For Horizon-AGN, there is a large-scale 1% excess of the $P_{\text{baryon}}/P_{\text{DMO}}$ ratio above unity that causes a relatively large change in the HZPT R_{nh} parameters (the source of this deviation is discussed in Appendix A of [106], and may not be physical).
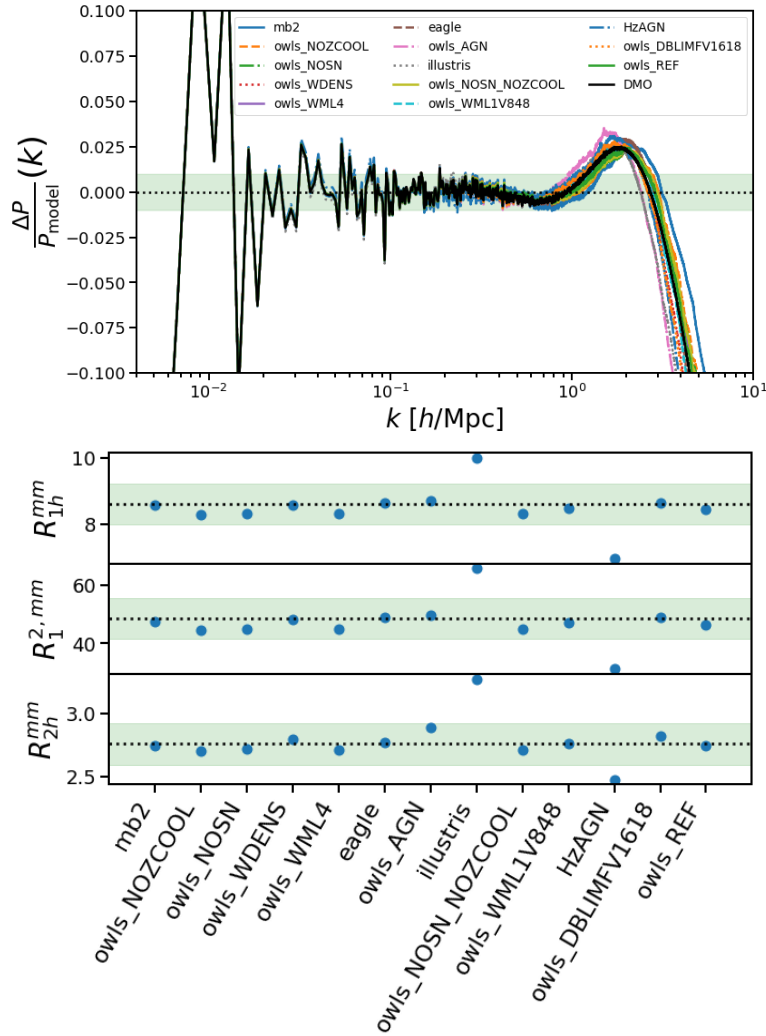
Figure 2.4: *Top:* Residuals for the power spectrum at $z = 0$ for all baryonic models considered. Shaded area corresponds to a 1% deviation. *Bottom:* HZPT parameter values fit for the different feedback models. DMO fits are denoted by the dotted lines and shaded areas denote rms deviations from the DMO case.

Extended power spectrum model

We briefly explore two extended models, which probe smaller scales where the effects of baryons are stronger, as modern cosmic shear measurements probe angular scales that receive contributions from these length scales. The first extension focuses on the one-halo term, as well as cosmology dependence of those parameters, and the second focuses on the two-halo term. We compute the power spectrum using `nbodykit` as described in Section 2.2 but with a finer mesh in the call to `FFTPower` with $N_{\mathrm{mesh}} = 10240$, and additional interlaced anti-aliasing [97]. This grid corresponds to a $k_{\mathrm{Nyq}} \approx 10 \ h/\mathrm{Mpc}$ and using the above settings should be trustworthy out to this scale [88].

Extending the one-halo term

We extend the model for $P_{mm}(k)$ to include one higher-order BB term ($n_{\max} = 3$) to get to 3%-level accuracy out to $k \approx 8 \ h/\mathrm{Mpc}$. We see an upturn in Figure 2.5 beginning at $k = 8 \ h/\mathrm{Mpc}$ which the model fails to fit. We see that the R_{nh} parameters can account for the strongest baryonic feedback (Illustris), which is perhaps not surprising given the fact that we have added two parameters - which are interpretable as the k^6 expansion coefficient reprocessed through the Padé expression.
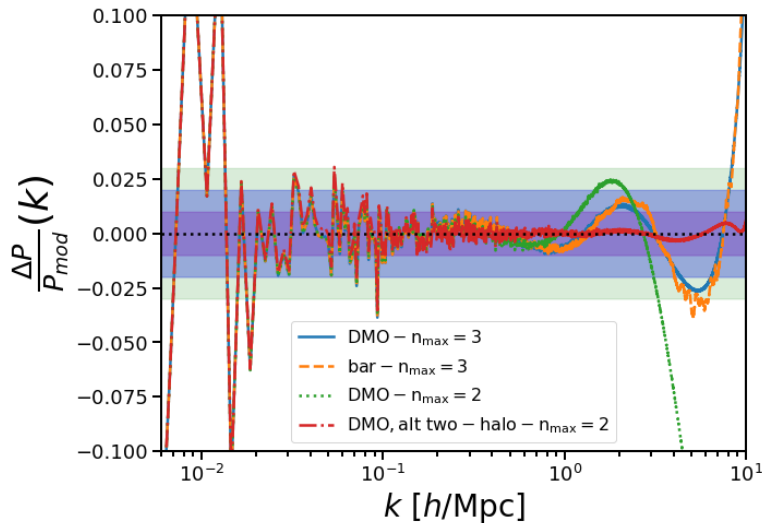
Figure 2.5: Fits to the matter power spectrum for dark-matter only in our simulations as well as with baryonic feedback effects included through the power spectrum ratio $P_{\mathrm{bar}}/P_{\mathrm{DMO}}$ for Illustris, the most extreme feedback model we consider, at $z = 0$ using the high-k extended model with $n_{\max} = 3$. We also show the alternate two-halo HZPT model described in Section 2.4 (red line). Colored regions are shown at 1, 2, and 3%.

Following SV15, we here provide fitted power-law dependence of all $n_{\max} = 3$ parameters for the $P_{mm}(k)$ (DMO) model described in the previous section at $z = 0$ (Eqn. 2.8). To

quickly obtain matter power spectra for cosmologies with different values of the matter density parameter Ω_m and the matter density variance in spheres of 8 Mpc/h σ_8, we use `CosmicEmu` [66], using 100 randomly generated power spectra with values of $\Omega_{cb} = \Omega_m^{\text{EMU}} \in [.26, .34]$ and $\sigma_8 \in [.7, .9]$ and 20 test spectra in the same range (Fig. 2.6). For the other emulator parameters we fix $h = 0.6774$, $\Omega_b = 0.0486$, $\Omega_\nu = 0.0014$, $n_s = 0.9667$ $w_0 = -1$, and $w_a = 0$

$$A_0 = 777 \left(\frac{\sigma_8}{0.8}\right)^{4.33} \left(\frac{\Omega_{cb}}{0.3}\right)^{-1.83} \tag{2.8}$$

$$R = 25.3 \left(\frac{\Omega_{cb}}{0.3}\right)^{-0.58} \qquad R_{1h} = 8.56 \left(\frac{\sigma_8}{0.8}\right)^{2.34} \left(\frac{\Omega_{cb}}{0.3}\right)^{-2.19}$$

$$R_1 = 7.34 \left(\frac{\sigma_8}{0.8}\right)^{2.37} \left(\frac{\Omega_{cb}}{0.3}\right)^{-1.39} \qquad R_{2h} = 2.93 \left(\frac{\sigma_8}{0.8}\right)^{1.56} \left(\frac{\Omega_{cb}}{0.3}\right)^{-1.24}$$

$$R_2 = 1.99 \left(\frac{\sigma_8}{0.8}\right)^{1.16} \left(\frac{\Omega_{cb}}{0.3}\right)^{-0.96} \qquad R_{3h} = 1.51 \left(\frac{\sigma_8}{0.8}\right)^{1.12} \left(\frac{\Omega_{cb}}{0.3}\right)^{-0.96}$$

The 3-parameter power law is accurate with rms residuals of 1% or less for the test set on all scales. In the context of this computation, this level of accuracy is competitive with state-of-the-art non-linear matter power spectrum models [119] - we provide a comparison to the right panel of Fig. 2.6 in Appendix 2.8. The few large-k residuals that go slightly past 2% correspond to the most extreme values of Ω_m^{EMU} at the edge of our range. We find positive exponents for all parameters with respect to σ_8 and generally negative ones for the parameters with respect to Ω_m^{EMU}. We note that fitted value for the exponent on the compensation scale R is close to zero ($< 10^{-15}$), so the value of R is essentially independent of the value of σ_8, and so we treat R only as a function of Ω_m^{EMU}.

Alternate two-halo term

As mentioned in Section 2.3, one might consider alternatives to ZA for the two-halo term in the HZPT model, and we briefly explore such an extension here[3]. One such alternative is based on the power due to the linear correlation function shifted by the ZA displacement:

$$P_{\text{alt}}(\mathbf{k}) = 4\pi \int dq q^2 \xi_L(q) \mathrm{e}^{-k_i k_j A_{ij}(q)} \tag{2.9}$$

where ξ_L is the linear correlation function and $A_{ij}(q)$ is the LPT displacement difference cumulant [43, 120, 17, 45, 121]. A benefit of taking P_{alt} as the starting point for a new two-halo term is that it remains easy to compute while preserving many of the advantages of ZA, and also provides slightly more power than ZA on quasi-linear scales. This modified

---

[3]We are grateful to Zvonimir Vlah for suggesting the main idea of this section.
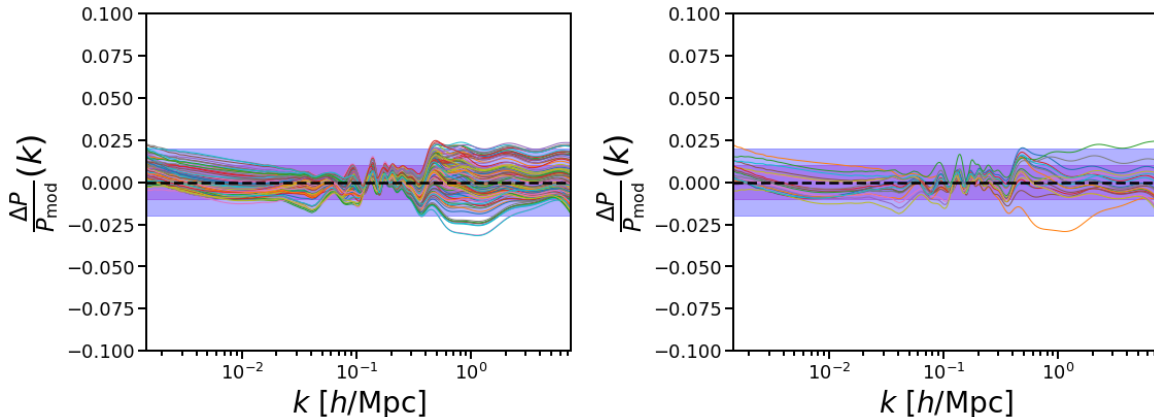
Figure 2.6: *Left:* Residuals of power law fits in the training set of 100 values of σ_8 and $\Omega_m^{\mathrm{EMU}} = \Omega_{cb}$. *Right:* Residuals of power law fits in the test set of 20 values of σ_8 and Ω_m^{EMU}. Colored bands show 1 and 2 %. See Fig. 2.17 for a comparison to `HMCode2020`.

version of ZA can be further augmented in a manner similar to the models of [122, 45] by adding a transfer function style polynomial term multiplying P_{alt}, such that

$$P_{\mathrm{alt,HZPT}}(k) = \left(1 + \alpha k^2 + \beta k^4\right) P_{\mathrm{alt}}(k) + P_{BB}(k) \qquad (2.10)$$

where α, β are free parameters. This model keeps with the HZPT spirit of an analytic Fourier transform and provides a converged inverse Fourier transform due to the Gaussian term of the ZA-like term[4]. For $n_{\mathrm{max}} = 2$, this model has the same number of parameters as the $n_{\mathrm{max}} = 3$ model described in Section 2.4. The success of this model in fitting the `CrowCanyon` matter power spectrum is demonstrated by the red line in Fig. 2.5. Clearly for the same number of parameters, $P_{\mathrm{alt,HZPT}}$ outperforms the $n_{\mathrm{max}} = 3$ model, achieving 1% residuals up to $k = 10\ h/\mathrm{Mpc}$. We also find that fixing $\beta = 0$ results in similar performance to the $n_{\mathrm{max}} = 3$ model, with one fewer parameter. This illustrates the power of the form of the HZPT model and that it is possible to further improve beyond the models we present in this paper within the HZPT framework by more carefully balancing the work sharing between the HZPT two-halo and one-halo terms.

We anticipate that using the form of eqn. 2.10 may also improve the accuracy of tracer HZPT models, and also results in exact expression for tracers using linear bias. We find that since for tracers we mostly limit our attention to wavenumbers below $1 - 2\ h/\mathrm{Mpc}$, using ZA alone suffices for our purposes. It would be quite interesting to further explore HZPT models of tracers based on P_{alt} or other similar two-halo terms.

---

[4] We find that fits to ξ^{mm} using $\xi_{\mathrm{alt}} + \xi_{BB}$ with $n_{\mathrm{max}} = 2$ are also accurate at the percent level.

2.5 Halos and Exclusion

In this section we present the HZPT model for halo-halo and halo-matter two-point statistics in configuration and Fourier space. While these quantities are not directly observable, understanding them is key to accurately modeling non-perturbative effects in the transition regime - namely, the discreteness of the halo field, and halo exclusion.

Small-scale halo clustering

The HZPT model was successfully applied to halo-matter cross-correlation two-point statistics, but was not applied to halo clustering (auto-correlation) statistics. H17 showed that fits for the power spectrum and correlation function to simulations were accurate to 2% to $k = 1$ h/Mpc (r $\sim$ 4 Mpc/h) between $z = 0 - 1$ for several halo mass bins between $10^{13} - 10^{14} M_\odot$. The model H17 used was the base HZPT model for dark matter with $n_{\max} = 2$ and with an additional linear bias parameter b_1. The HZPT parameters were fixed as fitted power laws of the linear halo bias b_1 and σ_8. Allowing all parameters (including b_1) to be free, we find that we can produce fits for these halo-matter correlators that are slightly more accurate on large scales corresponding to those of H17 (our bins 6-8) and also for a wider range of halo-mass bins (see Table 2.2) as shown in the right panels of Figs. 2.7 and 2.8. With the R_{nh} free we are able to fit down to smaller scales that H17 struggled with in modeling the correlation function. H17 attributed this failure of the configuration-space model to one-halo effects since the dependence on profiles cannot be fully described with a simple power law in b_1, and suggested a more complete treatment of nonlinear/nonlocal bias. We examine where the model fails and succeeds in more detail in the next section, but first turn to the non-perturbative features present in halo auto-correlations.

Non-perturbative modeling is necessary to accurately capture small-scale halo clustering. Ref [123] (B13) conducted a detailed study of halo exclusion and halo auto-correlation stochasticity using N-body simulations. Manufacturing a discrete halo field from the continuous matter field introduces Poisson noise from the finite number of resulting objects in a given volume. This contributes at zero-lag in configuration space and on all scales in Fourier space as the well-known "Poisson shot noise" or fiducial stochasticity. In the $k \to 0$ limit, however, because the tracer field is discrete, there is a constant contribution to the power spectrum that involves an integral over the correlation function. This means that the constant noise on all scales in Fourier space may be sub- or super-Poisson. The sub/super-Poisson noise has been investigated in detail [124, 125, 126, 123].

In addition to the scale-independent contribution from discretizing the field, the phenomenon of halo exclusion introduces a scale-dependent contribution to the two-point statistics. Halo exclusion follows directly from the foundational assumption of the halo model - that all matter is contained in non-overlapping collapsed dark matter halos. If halos are idealized as spherical, the phenomenon of exclusion appears quite straightforward. Since halos may not overlap, it is not possible for the (spherical) halo field to be correlated on scales below the sum of halo radii. This is reflected in a discontinuous drop to a value of -1

in the halo correlation function at the exclusion scale. For this simplified case of spherical halos at fixed mass, we can write the following expressions of B13 for the discrete (auto) correlation function for halos:

$$\xi_{hh}^{(d)}(r) - \frac{1}{\bar{n}}\delta^{(D)}(\mathbf{r}) = \begin{cases} -1 & r < R_{\text{exc}} \\ \xi_{hh}^{(c)}(r) & r \geq R_{\text{exc}}, \end{cases} \tag{2.11}$$

or, writing the two-point function, we have $\xi_{hh}^{(d)}(r) - \frac{1}{\bar{n}}\delta^{(D)}(\mathbf{r}) = (\xi_{hh}^{(c)}(r) + 1)\Theta_H(r - R) - 1$, where $R_{\text{exc}} = R_{\text{exc}}(M)$ is the exclusion scale, and in the notation of B13, (d) signifies "discrete" as to be distinguished from (c) "continuous"[5].

For the power spectrum the corresponding expression is:

$$P_{hh}^{(d)}(k) = \frac{1}{\bar{n}} + P_{hh}^{(c)}(k) - V_{\text{excl}}\left(W_R(k) + \left[W_R * P_{hh}^{(c)}\right](k)\right), \tag{2.12}$$

where $W_R(k)$ is the spherical top-hat window in Fourier space and $*$ is the convolution operator. We will consider these easily-interpretable toy expressions as conceptual references in a somewhat more realistic models of exclusion. In these models we introduce the exclusion scale R_{exc} as a free parameter.

In reality halos are not spherical, and even for a fixed-mass sampling of the halo field the scale at which exclusion sets in (the effective "exclusion radius") must necessarily reflect the fact that triaxiality leads to a distribution of "true" exclusion scales. However, if one interprets halos in the context of Lagrangian density peaks, then based on the 1-D findings of [127] (where no triaxiality can be present) peak exclusion is dominated by dependence on peak height, bin width, and peak curvature. These results appear to hold in 3-D as well [128], so the effect of triaxiality on exclusion is likely subdominant. Similarly, [129] found that triaxiality, substructure, and concentration scatter were negligible in modeling exclusion in halo two-point correlation functions. We also find it unnecessary to model these effects for percent-level accuracy. The criterion used to define the halo also has an impact on exclusion, which we return to in Section 2.6 and Appendix 2.10.

Correlation function results

In Fig. 2.7, we show the correlation functions from simulations (black points for halo-matter, gray points for halo-halo) as well as various HZPT models for different mass bins (see Appendix 2.9, Table 2.2) at $z = 0.55$. The linearly-biased ZA (black curves) agrees well with the simulations on the largest scales considered here, but significantly underestimates the correlation function at the several percent starting at $r = 40 - 60$ Mpc/h.

The second row of Figure 2.7 illustrates this deviation from linearly biased Zeldovich, which is fit by the BB terms. There are at least three scales in the enhancement over

---

[5]Here and in the remainder of this work, we take $\xi^{(c)}$ to contain any nonlinear or non-perturbative clustering outside the exclusion scale.

Zeldovich in both the halo-matter and halo-halo correction functions that the BB terms must fit to account for all halo masses considered here. These scales are 1. a large-scale enhancement (LSE) at ~ 10 Mpc/h (corresponding to the $n_{\text{max}} = 1$ parameter) 2. a small-scale enhancement (SSE) outside the halo exclusion scale (corresponding to the $n_{\text{max}} = 2$ parameter) and 3. the halo exclusion scale. These scales are clearly visible in the second row of panels in both halo-matter and halo-halo (though are more easily seen in halo-halo). We will first describe how these scales vary with halo mass (as seen in row 2 of Fig. 2.7) and then will describe how the HZPT models explicitly account for these scales.

The enhancement in the correlation function over ZA becomes more complicated for lower mass halos. For the largest halos ($M > 10^{13.5}$ $M_\odot/h$, right two columns), there is only one scale or "bump" visible in the enhancement - the LSE and SSE coincide at several Mpc/h. Just below this unified scale is the exclusion scale, which presents itself as a vertiginous climb to profile-dominated scales in the halo-matter CF, and as a precipitous drop to zero correlation in the halo-halo CF. For smaller halos, the single scale splits into the LSE and SSE, which are clearly visible at $\sim 5 - 10$ Mpc/h and $\sim 1 - 2$ Mpc/h, respectively, for halos with $M < 10^{13}$ $M_\odot/h$. Physically, the SSE may be connected to the non-perturbative enhancement outside the exclusion scale observed in peak clustering observed by Ref. [128], while the LSE may be more related to nonlinear bias (e.g. [130, 79]). It would be interesting to consider an expanded hybrid modeling approach in which the LSE is modeled with a more complex nonlinear biasing model than linearly-biased ZA as the two-halo term.

We reproduce the result of H17 - halo-matter CFs are well-fit by the $n_{\text{max}} = 2$ model above scales up to a few times the exclusion scale for all halo masses (green curve). This is because the $n_{\text{max}} = 2$ model captures two scales - the LSE and SSE - quite well, and in the larger halo mass case the values of the R_{nh} parameters increase and become much closer to each other, reflecting the unification of the LSE and SSE. We attempt to slightly improve upon the $n_{\text{max}} = 2$ model by adding a term to account for the halo profile dominance near the exclusion scale by adding a second BB term with $n_{\text{max}} = 1$ (purple dashed curve) without compensation [6]. This model does not show dramatic improvement over the $n_{\text{max}} = 2$ model, but does fit the outer portion of the profile dominated region quite well. There are small deviations in the halo-matter correlation function just outside the exclusion scale for both models - we speculate that these are the result of a too-simple treatment of halo compensation.

The halo-halo correlation function is well modeled for large masses by accounting only for the equal LSE-SSE scale and the exclusion scale through a modified $n_{\text{max}} = 1$ model. To model the step in the halo (auto)correlators we model the exclusion step in a similar manner to B13 using the function $F_{\text{exc}}(r)$

$$1 + \xi_{hh}^{(d)}(r) = F_{\text{exc}}(r) \left[1 + \xi_{hh}^{(c)}(r) \right], \tag{2.13}$$

---

[6]We set the compensation parameter of this additional term to be very large, 10^9, which is effectively the same as ignoring it.

where $\xi_{hh}^{(c)} = b_1 \, \xi_{\mathrm{hzpt}}$. The function $F_{\mathrm{exc}}(r)$ is an approximation to a more complete physical model for exclusion [127, 128] (we further discuss choices for this model in Appendix 2.9), and also return to it in the context of the power spectrum in Section 2.5. This model works quite well for the largest two halo mass bins (blue curve), but begins to fail dramatically when the LSE and SSE diverge at lower halo mass. To address this, we take the same strategy as for the halo-matter CF and upgrade the BB term to $n_{\mathrm{max}} = 2$ (red dashed curve). This results in excellent fits on scales for all halo masses in the halo-halo CF.

Power spectrum results

In Fourier space, the $n_{\mathrm{max}} = 2$ model is sufficient to capture the halo-matter cross power spectrum for all halo mass bins at 2% accuracy to $k = 1 \, h/\mathrm{Mpc}$, and at 1% for almost all mass bins on the same scales (Fig. 2.8). For the halo-halo power spectrum, the $n_{\mathrm{max}} = 1$ HZPT model in Fourier space with an added effective (not necessarily Poisson) shot noise term $\frac{1}{\bar{n}}$ appears to be accurate (thin lines in bottom panel of Fig. 2.8). In this scenario, R_{1h} should be thought of as a more general k^2 term rather than as a moment of the halo profile as is the case for matter. We discuss the small effect of removing R from the auto-correlation model in Appendix 2.9.

The $n_{\mathrm{max}} = 1$ model does not explicitly account for the Fourier space effects of exclusion, but fits quite well down to $k = 0.9 - 1 \, h/\mathrm{Mpc}$ at the 2%. For all halo masses, the scale-dependent correction due to exclusion is sub-dominant to the constant shot noise, (as seen in the left panel of Fig. 2.8), and for lower halo masses the model fails to be accurate at 1-2% at slightly lower k. This seems consistent with the results of [78], who are able to fit $P_{hh}(k)$ to lower maximum k using a Lagrangian bias expansion, including a k^2 term. However, exclusion must be properly modeled for percent-level accuracy in both configuration and Fourier space, and we return to this point in the following section, where we provide context for interpretation of the quoted accuracies in configuration and Fourier space with regard to exclusion.

As in the configuration space picture, modeling exclusion (using the Exp model presented in 2.9) as well as the LSE and SSE through the $n_{\mathrm{max}} = 2$ BB term extends the range of scales accessible to the power spectrum model. Proper exclusion modeling suppresses the observed deviations in the (thick) residuals to 1% below $k = 2 \, h/\mathrm{Mpc}$ and eliminates the need for a free constant shot noise parameter. In the fits shown in the thick residuals in the bottom panel of Fig. 2.8, the shot noise is fixed to Poisson, and the correction comes entirely from the exclusion model. Using the exclusion model with an $n_{\mathrm{max}} = 1$ BB term suffices at the same level of accuracy for the highest two mass bins, but (as discussed in 2.5) for the lower mass bins the results are worse since the SSE and LSE are distinct and must be modeled separately by something more flexible than the $n_{\mathrm{max}} = 1$ BB term. While we only go to $k = 2 \, h/\mathrm{Mpc}$ here, the Fourier space exclusion model is accurate to even smaller scales (see Section 2.5).
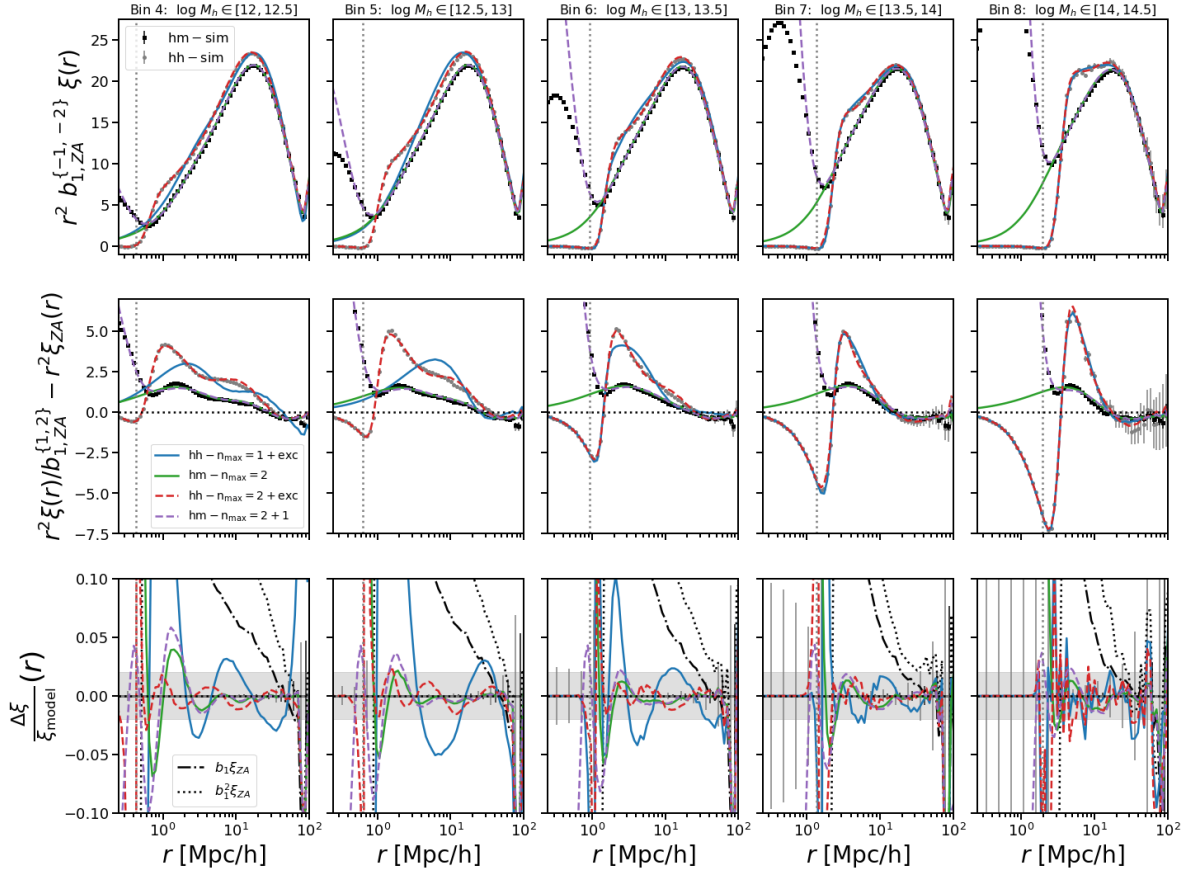
Figure 2.7: Fits to the halo-halo and halo-matter correlation functions in several logarithmic mass bins. The top panels show the halo-halo and halo-matter correlation functions (multiplied by r^2 and divided by the b_1^2 and b_1, respectively), the center panels isolate the correction to ZA by subtracting it out, and the bottom panels show residuals with a shaded band at 2%. The columns correspond to increasing halo mass from left to right. Green (purple dashed) lines show the $n_{\max} = 2$ ($n_{\max} = 2 + 1$) HZPT halo-matter correlation function, while blue (red dashed) lines show the HZPT $n_{\max} = 1$ ($n_{\max} = 2$) halo-halo correlation function. Black points show the halo-halo simulation correlation function, and grey points the halo-matter simulation correlation function. Errors are Fourier transformed diagonal Gaussian+Poisson, which are meant as a visual guide only as errors are correlated. The number of residual points with errorbars in the bottom panel has been reduced for visibility. Vertical dotted lines mark the minimum scale used to fit each mass bin, which is roughly the lower limit of the transition regime for each mass bin ($\sim R_{\mathrm{Lag}}/2$).

Transforming the two-point statistics

Exclusion and non-perturbative clustering present themselves differently in configuration and Fourier space. To better understand how to interpret the accuracy of the HZPT model
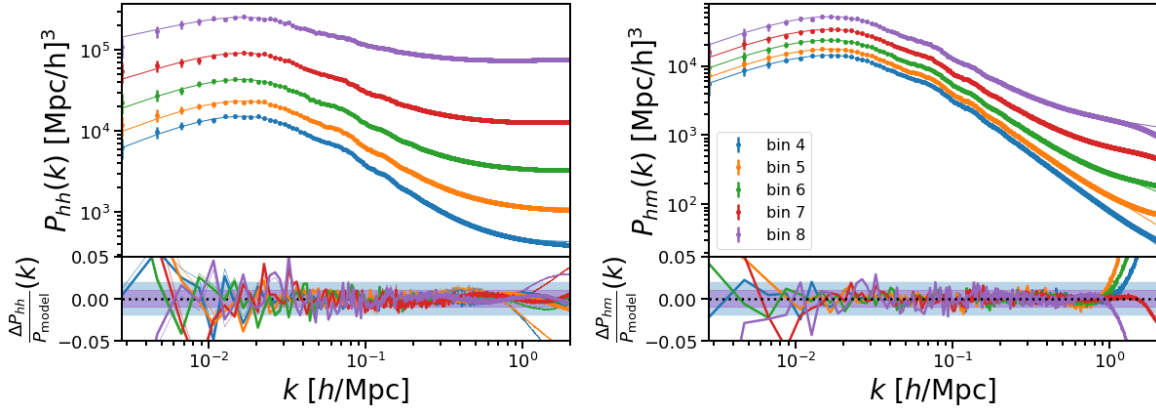
Figure 2.8: *Left:* Halo-halo power spectra residuals from and the best fit HZPT models at different halo masses. Thin lines are residuals for the $n_{max} = 1$ model with a free shot noise parameter, while thick lines are due to the model of exclusion with $n_{max} = 2$ similar to the best model in Fig. 2.7. *Right:* Fits for halo-matter power spectra using the $n_{max} = 2$ model. Colored bands are shown at 1 % and 2 %.

at the different scale cuts in k and r, we provide a narrow comparison of models that include and exclude the non-perturbative effects of exclusion in configuration and Fourier space for a single halo mass bin (bin 7 - though the result for other bins is similar[7]) in Fig. 2.9.

In the left panel of Fig. 2.9, we show two realistic models for the exclusion step in configuration space - the ErfLog model (solid orange) is the model of B13 (with free σ_{exc}, R_{exc}, two free parameters total), and the qualitatively similar Exp model (one free parameter R_{exc}, green dash-dotted) are both described in Appendix 2.9. The Exp model fits the step quite well, though the shape is not quite right at the smallest scales of the step, and the ErfLog model clearly captures the exclusion step even better than the Exp model. The ErfLog model is relatively insensitive to σ_{exc} - fixing $\sigma_{exc} = 0.1$ (similar to B13) only mildly degrades the accuracy of $\xi_{hh}(r)$ in the $n_{max} = 1$ model (which is sufficient here for bin 7) with respect to the shape at the small-scale end (similar to the slight inaccuracy of the Exp model). A benefit of the Exp form of the exclusion step is that it permits an analytic Fourier transform of $F_{exp}(r)$ (see Appendix 2.9), which explicitly displays the non-trivial k-dependence of the exclusion (it is not as simple as k^2) and keeps with the spirit of the HZPT model. Clearly both of these models are capturing the correct features of non-perturbative halo clustering on the smallest scales.

For comparison, we also show the TH model (blue dashed), which is given by the simple Heaviside truncation of the $n_{max} = 1$ HZPT model fit down to $r = 2.5$ Mpc/h, as well as the same model fitted using a larger minimum scale $r = 10$ Mpc/h (solid purple), which we refer to as the "ls" (large-scale) model. Similar to the quadratic biasing model of B13,

---

[7]However, for lower mass bins, the large-scale finite-size correction is super-Poisson rather than sub-Poisson, which is expected from the explanation of B13.

the ls model describes the LSE but fails to capture both the full extent of the SSE outside the exclusion scale, and totally misses the exclusion step. The un-truncated TH model ($\xi^{(c)}$) correctly describes the SSE but fails to account for the exclusion step - by adding the Heaviside truncation, the TH model provides a qualitatively correct description of both the SSE and the exclusion step. However, quantitatively the TH model does somewhat worse than the ErfLog and Exp models by failing to account for the finite width of the step.
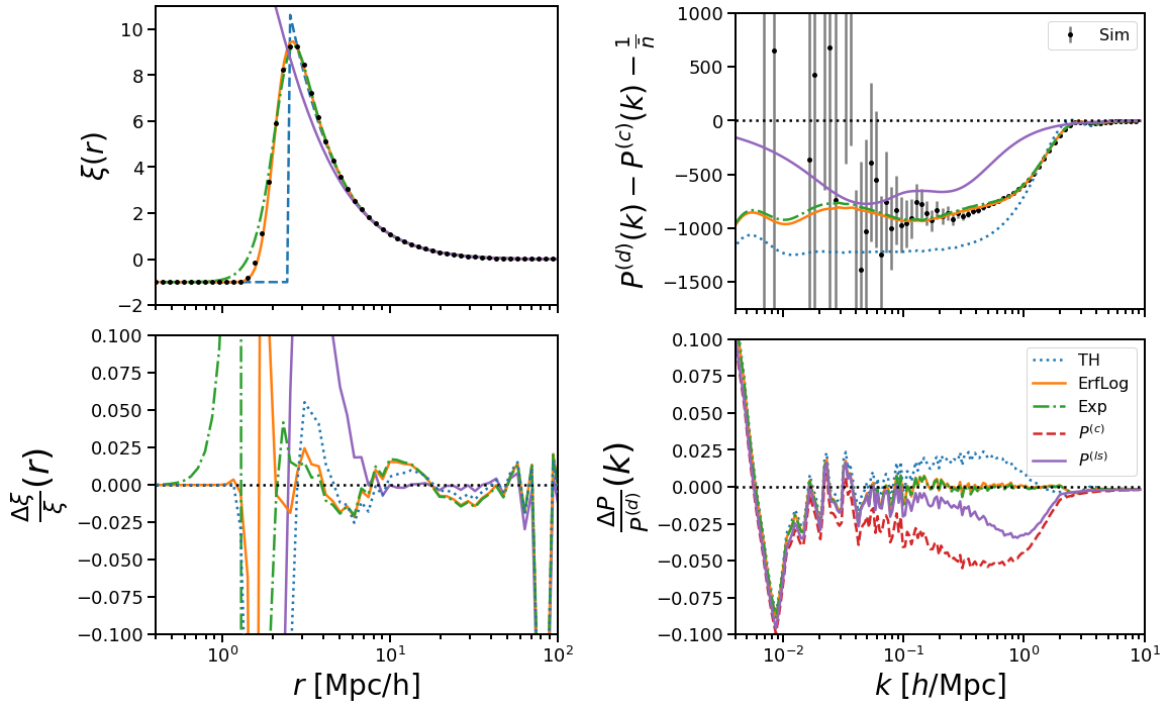
Figure 2.9: Illustration of (HZPT) contributions to exclusion for bin 7 ($\log M_h \in [13.5, 14]$. *Left:* Correlation function with several different exclusion models. ErfLog denotes the model of B13 (with σ_{exc} free), while Exp denotes the model described in Appendix 2.9 (that has correction with analytic FT). TH denotes the simple thresholding of the continuous model, and is clearly not sufficient to capture the width of the exclusion step. *Right:* Corresponding power spectra (evaluated via FFTLog). $P^{(c)}$ denotes the (analytic) continuous model using best-fit HZPT parameters to the correlation function. TH in this case is equivalent to the window function expression terms of B13 (but without truncating the expansion). The solid purple curve denotes the large-scale model $P^{(ls)}$ resulting from fitting $\xi^{(ls)}$ above 10 Mpc/h.

Fourier transforming, we look to see what happens under the various treatments of exclusion in configuration space. The result is shown in the right panel of Fig. 2.9. Here the quantity on the vertical axis (top panel) is a measure of the error induced by not treating the exclusion step and simply using the "continuous model" $\xi^{(c)}$ with fiducial (Poisson) shot

noise $\frac{1}{\bar{n}}$. $P^{(c)}(k)$ is the (analytic) Fourier transform of $\xi^{(c)}(r)$ that is fit up to the peak of the halo correlation function (TH in the top left panel but without the threshold). The simulation points in the top-right panel and the red curve in the lower-right panel illustrate that modeling the SSE but totally ignoring the exclusion step induces a 5% error on scales between $k = 0.1 - 1\ h/$Mpc, which is actually larger than if we had simply used $P^{(ls)}$, the Fourier transform of large-scale model $\xi^{(ls)}$. This can be understood from the fact that the non-Poisson correction in the large-scale limit ($k \to 0$) is given by the integral $\int d^3 r \xi^d(r)$. The large-scale model $\xi^{(ls)}$ both underestimates $\xi^{(d)}$ on scales where the SSE is relevant and overestimates it when the discrete correlation goes to zero, resulting in an accidental, but only partial, cancellation in the integral. This results in the wrong non-Poisson correction, but one that produces a smaller residual in the bottom-right panel Fig. 2.9 than for the continuous model $P^{(c)}$. $P^{(c)}$ only includes positive contributions to this integral, and produces a too super-Poisson correction without the negative contributions from the exclusion step. Adding the simple Heaviside threshold qualitatively accounts for both types of correction and tames the residuals to be about 2-3%. Using the more accurate exclusion step models produces negligible error at a fraction of a percent.

We note that these residuals are computed with respect to the discrete power spectrum (which includes fiducial shot noise), so all errors would be amplified were shot noise subtracted[8]. Since the effect of ignoring exclusion is scale-dependent, it is not possible to cleanly relate cutoffs at a particular scale in configuration space with cutoffs at a particular wavenumber k in Fourier space. Furthermore, we see through the comparison of the residuals of $P^{(ls)}$, $P^{(c)}$, and the step models that starting on quasi-linear scales and fitting a configuration space model down to progressively smaller scales induces an exclusion correction in Fourier space that initially increases before decreasing to approach the correct non-Poisson value. It would be interesting to further quantify how this behavior affects other biasing models that stop at a particular minimum scale $r_{\min}$ such that $R_{\rm exc} < r_{\min} < 10$ Mpc/h.

By way of this example, we see that ignoring the effects of exclusion (or stochasticity in general) as is sometimes done in small-scale biasing models of the tracer correlation functions, will necessarily result in incorrect behavior over a relatively wide range of wavenumbers in Fourier space. Coming back to our fits to the power spectrum, this is seen in Fig. 2.8, where percent-level excursions are visible in P_{hh}. These excursions seem more pronounced at lower mass, reflecting the fact that the distinct LSE and SSE identified in the correlation function are not well-modeled by a single $n_{\max} = 1$ BB term. The apparently smaller residuals for higher mass bins is, however, partially a consequence of shot noise constituting a larger fraction of the total power on small scales due to lower halo number density at higher mass.

Despite the fact that the $n_{\max} = 1$ fits in Fig. 2.8 do not explicitly include exclusion, they are still relatively accurate since the Fourier space BB term and free shot noise terms are implicitly modeling exclusion. This may be explained by the fact that at leading order the

---

[8]This amplification would of course get worse at higher k as the amplitude of the continuous power spectrum decreases - in practice quick calculations residuals with respect to the shot noise subtracted power spectrum for bin 6 can be above 10% at $k = 1\ h/$Mpc

$n_{\max} = 1$ BB term and the simple threshold model scale as k^2, so the BB term (or any k^2 term) might capture the leading-order behavior of exclusion. However, a more realistic model of exclusion (e.g. the Exp or ErfLog model) is more complicated than a simple k^2 term (see Appendix 2.9 for the form of this expression for the Exp model). These models introduce an additional parameter R_{exc} to the HZPT model, but knowledge of this parameter *exactly gives* the value of the non-Poisson correction in the large-scale limit, eliminating the need for the free shot noise parameter in the power spectrum. We conclude that percent-level accuracy of a model without exclusion in Fourier space is largely due to the relative importance of the free shot noise term and the leading-order k^2 behavior of exclusion, and *not* due to a correct model of high-k behavior, which must include the non-perturbative effects associated with halo exclusion.

2.6 Galaxies and Satellites

The transition regime for galaxy-galaxy and galaxy-matter correlators is affected by both the details of satellite occupation and halo exclusion. The model of Section 2.5 for halos already includes exclusion, and we build upon that model by accounting for the presence of satellite galaxies in this section. We consider two different galaxy samples produced according to the HOD prescriptions presented in Section 2.2 to test the flexibility of this model. We focus on configuration space fits and use them to estimate joint density with cosmology but also provide fits to power spectra. We take a pragmatic approach throughout this section, using the minimal HZPT model necessary to achieve percent-level accuracy for the galaxy two-point correlators at $k \approx 1$ h/Mpc and $r \approx 1 - 2$ Mpc/h.

Small-scale galaxy clustering

The galaxy-matter correlators are analogous to the case of halo-matter correlators, but are slightly complicated by the presence of satellites. On the largest scales considered here, galaxy-matter correlators are well-described by linear bias with ZA and compensation. In addition to the correlation between a particular central galaxy and the matter profile of its host halo, there is now another contribution from the correlation between matter and satellites. The satellite fraction will impact the amplitude of the intra-halo correlations, which in turn will affect the slope and location of the transition feature described in Section 2.5. The smallest scale correlations are then completely governed by the halo dark matter and satellite profiles. In our simple HOD mocks, satellites are drawn from an NFW profile. However, since the form of the BB term is profile-agnostic due to the general form of the Padé expression, there should be no great difficulty in modeling other qualitatively similar profiles (i.e. more complicated satellite profiles).

   The galaxy-galaxy correlation function ξ_{gg} is more complicated than the auto-correlation function for halos. In addition to the steep drop in the correlation function near the effective exclusion scale that is expected for halos, we must consider the role of satellite galaxies.

As explored in detail by [131] (see their Fig. 1) and H17, there are additional types of correlations: 1. between centrals and satellites in different halos, 2. and centrals and satellites in the same halo, as well as 3. between satellites in different halos, and 4. satellites in the same halo. While 2. and 4. essentially serve to change the correlation function on scales relevant to the satellite profiles (i.e. roughly the combination of the profile and its self-convolution), 1. and 3. effectively introduce contributions that are versions of the central-central correlation function (with the exclusion step) that have been smoothed out over the halo scale. This smoothing of the exclusion scale will serve to broaden the exclusion step present for halos in the galaxy auto-correlation.

To deal with these complications, we introduce some additions to the HZPT model. We do not model each of the terms outlined in [131] separately, instead lumping some of them together into an effective HZPT model for $\xi_{gg}(r)$. We allow for the smoothing of the exclusion step through freeing σ_{exc} to be larger than the value (≈ 0.1) that was acceptable for halos. Additionally, we add a satellite profile term (an additional BB term with $n_{\mathrm{max}} = 1$) that has two free parameters (with subscript $1s$), since we fix $R_{1s} = 10^3$. On these scales, this choice is the same as providing no compensation for the satellite profile. The A_{1s} and R_{1s} will vary depending on the details of the satellite occupation, e.g. with the amplitude scaling with the satellite fraction. So the full equation for galaxies with exclusion (in configuration space) is:

$$
\begin{aligned}
\xi_{gg}^{(\mathrm{exc})}(r) &= \xi_{hh}(r) + b_1 \xi_{BB,1s}^{n_{\mathrm{max}}=1}(r) \\
&= b_1 \left[F_{\mathrm{exc}} \left(\xi_{ZA} + \xi_{BB}^{n_{\mathrm{max}}=1} \right) + \xi_{BB,1s}^{n_{\mathrm{max}}=1}(A_{1s}, R_{1h,1s}) \right]
\end{aligned}
\tag{2.14}
$$

where the bias b_1 is free (not fixed to the halo bias value) and we suppressed arguments except for the new parameters in the second line. Here the 1s BB term is compatible with the usual BB interpretation and we can think of it (correctly) as a k^2 expansion in the satellite profile. Adding this term does not ruin the analytic Fourier transform, which will have a form that is the product of two Lorentzians (as for matter) in Fourier space.

The effects in the transition regime for HOD mock galaxy clustering will necessarily be more complicated than that of halos (even ignoring satellites) since the HOD applies a threshold for the central occupation which spans the equivalent of several halo mass bins. This means that the "cross-stochasticity" (B13, [124]) of exclusion in different bins will contribute more strongly to the central auto-correlations. However, we find that this is not something that needs to be modeled explicitly when fitting, which may have to do with the fact that the cross-stochasticity is either close to constant or of a similar scale dependence to that of the auto-stochasticity (c.f. H17 Fig 4). We also do not explicitly account the effects of central galaxy off-centering [132, 133], which are relevant for an accurate treatment of the small-scale galaxy-galaxy lensing signal. Since this effect may be accurately modeled by a modification of the profile, we anticipate that our Padé term may be general enough to account for such effects.

Two HOD mocks

The CM and LZ samples are produced using two different underlying simulations, and complement each other in the trade-off of resolution and number of simulated cosmologies. The CM sample is produced from a simulation with a factor of 10 larger volume than the LZ sample and allows for a cleaner test of model accuracy due to a reduction in resolution effects.[9] Fits to mocks from the LZ sample have increased noise with respect to to CM mocks due to smaller volume and a resulting smaller number of HOD galaxies, but still allow us to map the HOD basis of parameters onto the HZPT basis of parameters and provide a joint distribution of the HZPT parameters and cosmological parameters.

The CM sample covers a wide range of HOD parameters that are centered on the BOSS CMASS parameter space to illustrate the flexibility of the HZPT model (parameter space described in Section 2.2). The parameter space covers a wide range of satellite fractions ($f_s = 0.01 - 0.65$), the highest values of which are still consistent with observed galaxy samples [135] (though these may differ significantly from BOSS). We emphasize that this choice of "CMASS" parameters is not to be taken too literally, as we do not enforce that the HOD mocks reproduce the CMASS clustering, only approximately the CMASS number density and are produced at a similar redshift. The HOD parameter ranges are based on [82], but are taken to be more general - and use FoF rather than SO halos. The purpose of this sample is more illustrative and conceptual - to demonstrate that a wide range of HOD parameters can expose halo exclusion for certain mock galaxy samples, and that HZPT provides a good description of two-point correlators even in this case. The LZ sample is more realistic in the sense that the HOD parameters are close to those favored by LOWZ clustering [136, 137]. These differences mean that a different minimal HZPT model is necessary to reach percent-level accuracy for each sample at the scales we address in this section.

Configuration space results

The galaxy-matter cross correlators in both the CM and LZ mock samples are well fit by the same model used for halo-matter cross correlation - the base HZPT model with $n_{max} = 2$ with the linear bias b_1. Despite the presence of satellites, the transition regime (including the outer part of the halo profile for CM) is well-modeled by the R_{nh} parameters. Fits are performed from $r = 1 - 70$ h/Mpc in configuration space (and from $k = 0.01 - 1$ h/Mpc in Fourier space). Errors for fits to the LZ power spectra include diagonal Gaussian+Poisson covariance and variance estimated from repopulating HODs at 10 different random seeds (which we take as independent of any particular HOD realization), while the CM power spectrum errors are based only on Gaussian+Poisson covariance.

---

[9]The qualitative features of the exclusion step do not appear to depend much on the use of FastPM. Using a subset of the same HOD parameters (albeit at a slightly different cosmology) we check the qualitative features (and scales) of the exclusion feature are similar using FoF catalogs produced by a TreePM code ([134], described in H17).

The $n_{\max} = 2$ HZPT model is sufficient to attain several-percent accuracy in both the galaxy-matter and galaxy-galaxy correlation function (generally 1-2%, but this is limited by uncertainty due to resolution in the case of LZ - the rms error is always less than 1% [2%] below 40 Mpc/h for ξ_{gm} [ξ_{gg}]) down to $r = 2$ Mpc/h for LZ. For CM the $n_{\max} = 2$ model is also sufficient for 2%-level accuracy ξ_{gm} $r = 1$ Mpc/h, but for ξ_{gg} we require the model of eqn. 2.14 to capture the complications due to the satellites and exclusion effects present in this sample to produce an accuracy of 2% above $r = 2$ Mpc/h. We find that for most choices of HOD parameters, fits in both gm and gg (with the satellite terms) provide fits accurate to 1% down to $r \approx 0.5$ Mpc/h for the CM sample. However, to be conservative and accommodate all HOD parameters considered, here we fit ξ_{gg} only down to $r = 2$ Mpc/h. The galaxy-matter correlation function ξ_{gm} is fit to 1 % accuracy down to $r = 1$ Mpc/h for all but three of the highest satellite fraction HODs, in which case the accuracy quoted is 2% (Fig. 2.10). All CM mock HODs with $f_s > 0.55$ (very high for realistic LRG samples even given the general parameterization used here) have correlators plotted as gray curves in the figures. There is a downward shift in scale in the transition regime between the two samples, which we discuss in Section 2.6.
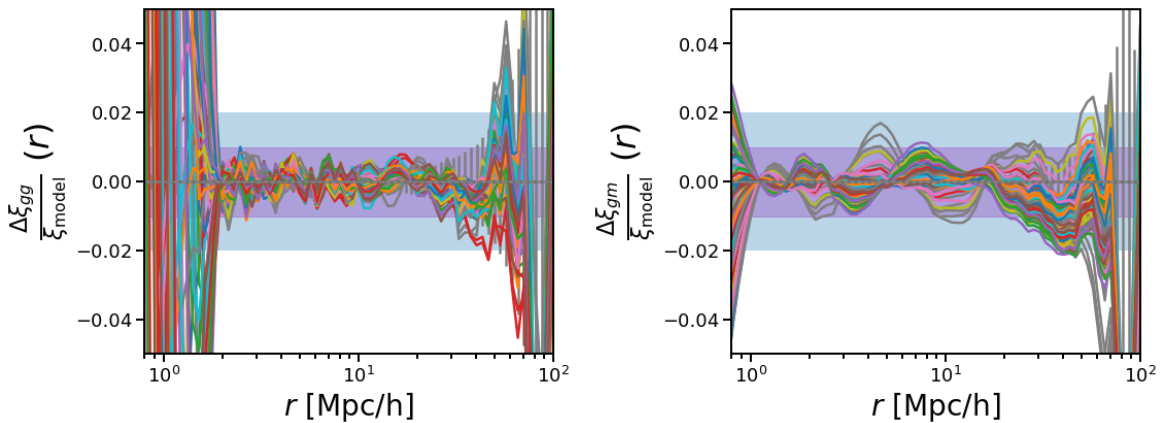
Figure 2.10: *Left*: Galaxy-galaxy correlation function residuals for the 100 CM HOD mocks. *Right*: Galaxy-matter correlation function residuals for the 100 CM HOD mocks. Grey curves correspond to HOD mocks with $f_s > 0.55$. Colored regions mark 1% (red), and 2% (blue), errors are diagonal Gaussian+Poisson (these are meant as a visual guide only, and the errors are correlated). The HZPT model used in these fits is the $n_{\max} = 2$ model for ξ_{gm} and the satellite-enhanced $n_{\max} = 1$ model with exclusion (eqn. 2.14) for ξ_{gg}.

Fourier space results

As discussed in Section 2.5, exclusion in Fourier space is largely suppressed by scale-independent shot noise, and the $n_{\max} = 1$ model performs decently well for $P_{gg}(k)$. We see that for all but three of the HOD mock power spectra considered, the residuals are always less than 3%,
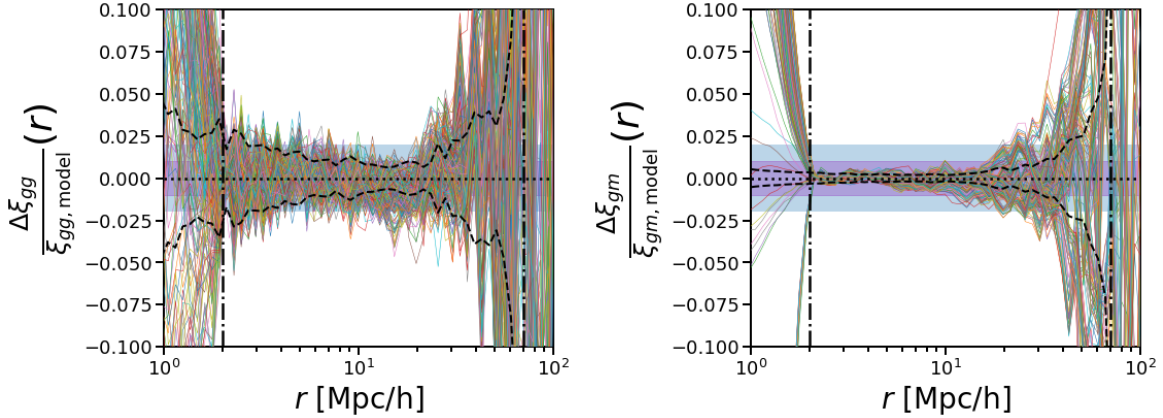
Figure 2.11: *Left:* Residuals for the $n_{\max} = 2$ model for ξ_{gg} for all 200 cosmology and HOD parameter combinations. (*Right:*) The same for ξ_{gm}. Colored bands show 1 and 2%, and errors (black dashed) give 1σ Gaussian+Poisson variance (these are meant as a visual guide only, and the errors are correlated), with additional stochastic HOD error added in quadrature to the ξ_{gg} error. The HZPT model used in these fits is the $n_{\max} = 2$ model for both ξ_{gm} and ξ_{gg}. The more involved model of eqn. 2.14 is not necessary for percent-level accuracy for the LZ sample.

and are usually less than 1%. Again, gray curves correspond to HOD mocks with $f_s > 0.55$, and account for the residuals exceeding 2%. The three offending curves correspond to the $f_s > 0.5$ cases mentioned for the correlation function. From the top panel of Fig. 2.12, we can see that these are the highest-biased cases and the wiggle-shape of the excursions in the transition regime are what would be expected from ignoring exclusion, as we have done here. Contrary to the case of halos, the number density is fixed in the CM sample, so the fiducial shot noise is kept fixed as the preference for population of galaxies in halo masses changes. For the highest mass halos, we would not expect to see the transition wiggle feature due to the high value of shot noise, but for these high satellite fraction models high-mass halos are preferred and we essentially reduce to the case of a high-mass halo bin where the exclusion feature is smoothed and shot noise is reduced (satellites act to up-weight the importance of the halo correlation with respect to the shot noise).

The fits in the power spectrum should not be taken to mean that exclusion is not important for an accurate description of two-point statistics, and the lessons of mapping between configuration space and Fourier space recounted in 2.5 still apply.

We find that the $n_{\max} = 2$ HZPT model used for the galaxy-matter correlators is sufficient to attain several-percent accuracy (generally 2-3%, as limited by uncertainty due to resolution - the rms error is always less than 2% over the range of fit) down to $k = 1h/\mathrm{Mpc}$ for the galaxy auto-correlators for the LZ sample. For the LZ HOD mocks, the exclusion feature is essentially undetected by the fits on the scales we consider (and the feature is totally absent for 90% of the HOD mock correlation functions) and is not present by eye in Fig. 2.11.
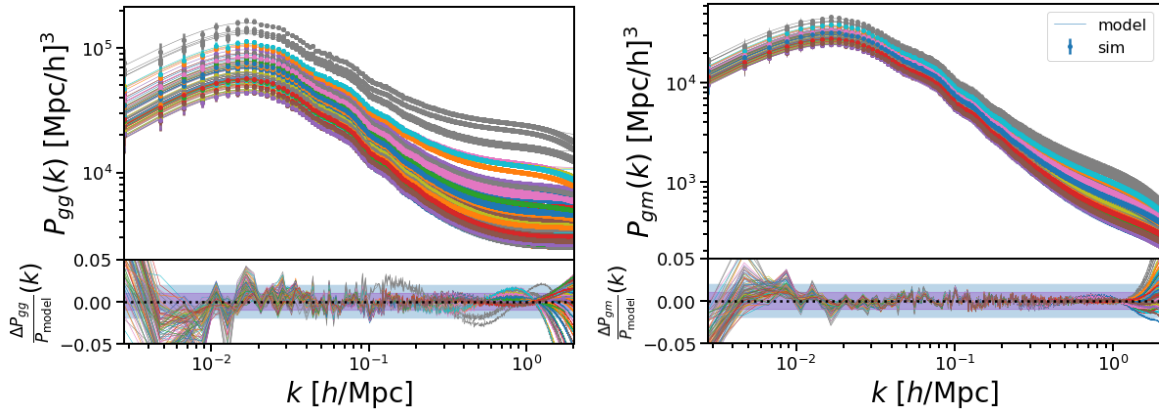
Figure 2.12: *Left*: Galaxy-galaxy power spectra and residuals for the 100 CM HOD mocks. *Right*: Galaxy-matter power spectra and residuals for the 100 CM HOD mocks. Grey curves correspond to HOD mocks with $f_s > 0.55$. Colored regions mark 1% (red), and 2% (blue). The HZPT model used in these fits is the $n_{\max} = 2$ model for P_{gm} and the $n_{\max} = 1$ model without exclusion for P_{gg}.
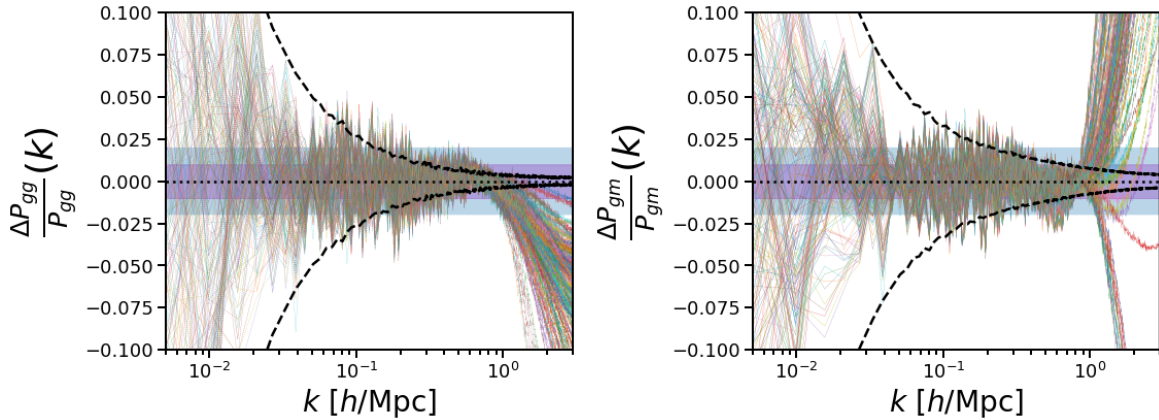
Figure 2.13: *Left:* Residuals for the $n_{\max} = 1$ model for P_{gg} for all 200 cosmology and HOD parameter combinations. *Right:* The same for the $n_{\max} = 2$ model (without exclusion) for P_{gm}. Colored bands show 1 and 2%, and errors (black dashed) give 1σ Gaussian+Poisson variance, with additional stochastic HOD error added in quadrature to the P_{gg} error.

Thus, the more complicated model including exclusion described above (and necessary for CM) is totally unnecessary. This is due to the position of the exclusion feature for the LZ sample, which is at particularly small scales, and which we discuss in the next section.

We find that using an augmented two-halo term similar to the one described in Section 2.4 can result in fits that are accurate at $k \approx 10 \, h/\mathrm{Mpc}$ with a single transfer function parameter

for P_{gm} in some HOD realizations. We also find that exclusion modeling in addition to a k-space version of the satellite profile term similar to eqn. 2.14 can result in improved fits in P_{gg}. An extension of the results presented in this section to higher k using an augmented two-halo term and exclusion is therefore an interesting direction for future work employing HZPT modeling.

Comparison of mocks

Exclusion is much more prevalent in the CM sample than in the LZ sample. When exclusion does occur, it does so on scales of $0.6 - 0.7$ Mpc/h in LZ[10] rather than around 2 Mpc/h as seems typical for CM. For $\log M \in [13.5, 14]$ (bin 7), which is completely covered by the selection of halos populated by the HOD mocks, the CM exclusion step spans $1.3-2.6$ Mpc/h, while for LZ it spans $0.7 - 1.0$ Mpc/h. This is consistent with the scales of the dips of the exclusion features in the galaxy-galaxy correlation function for both samples (Fig. 2.14). For the LZ mocks, exclusion is only visible for very low satellite fractions ($f_s < .05$), while for the CM sample it is present the majority of the time ($\approx 80\%$). For CM, in fact, the only time the model does not seem to show exclusion is when $\sigma_{\log M}$ is very large - suggesting that the sharp climb of the halo occupation may to some extent be driving the visibility the exclusion feature.

A reason for the difference between CM and LZ is that both the choices of halo finder and halo mass used strongly influence the exclusion feature. The halo catalog for LZ was produced by `ROCKSTAR` and uses "strict SO" 200b masses, which includes unbound particles that are not part of the `ROCKSTAR` group in a given halo. The CM halo catalog simply uses FoF (as implemented in `nbodykit`) masses with linking length $b = 0.2$. We describe this effect in further detail in Appendix 2.10.

Literature to date focusing on exclusion has relied on both FoF halos (e.g. B13, [128]) and SO halos (e.g. [129]). If one is consistent in a choice of FoF halos or SO halos in a mock-based analysis using HOD galaxies, the treatment of exclusion will also be consistent. The LZ implementation is thus more realistic in the sense that the analysis of [82] used a SO halo finder to obtain the HOD constraints. But even within the context of SO masses, the effect of SO vs strict SO masses can have an effect on the correlation function due to the rapid growth of the exclusion feature at small scales (Fig. 2.21). In HOD-based modeling it is then desirable to in some sense marginalize over halo definition. Ref. [136] used a free parameter (R_{rescale}) to do this, using the justification that the main effect of varying halo definition is the effect on halo radii as related to matter and satellite profiles. However, as shown here, the impact of halo definition on tracer correlators is not just through the one-halo term, but also through the two-halo term via exclusion.

While in Section 2.5, finding the minimum effective HZPT model that is accurate at the percent level required careful modeling of non-perturbative halo clustering, for mock HOD

---

[10]This comparison was performed using Box 20 of the Aemulus simulations, which is close to the cosmology used to generate CM, but for more disparate cosmologies the effect may be significant (already the fact that some boxes exhibit exclusion and some do not may point in this direction)

galaxies we find the reality to be more complicated. Depending on the galaxy sample and host halo definition, exclusion may or may not be an effect that is necessary to model at this level of accuracy. The CM sample contains HOD populations that often require a more involved model of halo exclusion and satellite contributions, while the LZ sample only shows exclusion on scales smaller than 1 Mpc/h, so we can safely ignore them when aiming for percent-level accuracy above these scales.
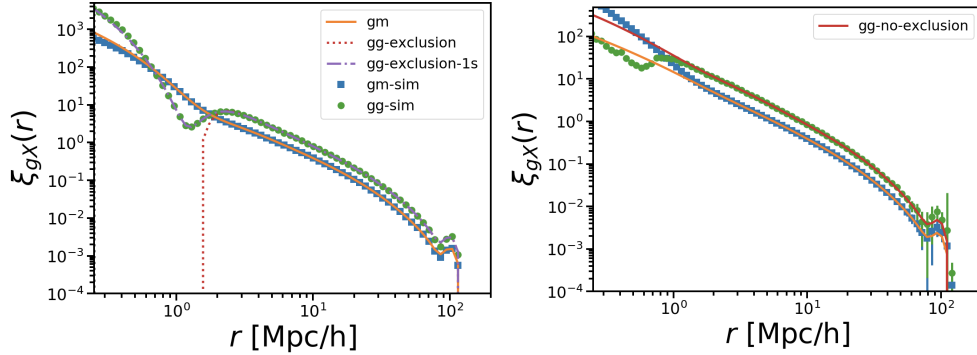
Figure 2.14: Comparison of fits to galaxy correlators for different samples when exclusion is present. Correlation function for mock galaxies for a single HOD from the CM sample (with satellite fraction $f_s = 0.29$). *Right:* Correlation function for mock galaxies for a single HOD from the LZ sample (with satellite fraction $f_s = 0.01$).

Correlations with cosmological parameters

We use the LZ sample and the best-fit HZPT parameters for galaxy-matter and galaxy-galaxy two-point correlators to outline an approach to estimating a joint prior of HZPT model parameters and cosmological parameters. In particular, we use Sliced Iterative Generator (SIG) [138] to perform density estimation using the best-fit HZPT parameters obtained from the set of 200 LZ correlation functions as training points. Potential analyses employing HZPT as the model for two-point correlators could then use such a density as a prior for analysis. Fig. 2.15 shows projections of a reduced density (considering only Ω_m and σ_8) for HZPT parameters that vary significantly with Ω_m and σ_8. The fitted R, R_1^2, R_{2h} parameters not shown do vary significantly with Ω_m and σ_8, but have strong degeneracies with R_{1h} so have been removed for visibility. We also note that, as mentioned in the previous section, we are usually not fitting non-perturbative effects near halo scales in the LZ sample.

This approach trades interpolation error (as in a standard emulator) for approximation error (through density estimation) and is more flexible than a typical emulator, as one can tweak the priors manually without running a new set of expensive simulations. Of course, there should be a physical reason for shifting or narrowing the priors, but simply broadening the priors may account for increased uncertainty about a particular galaxy or tracer sample.
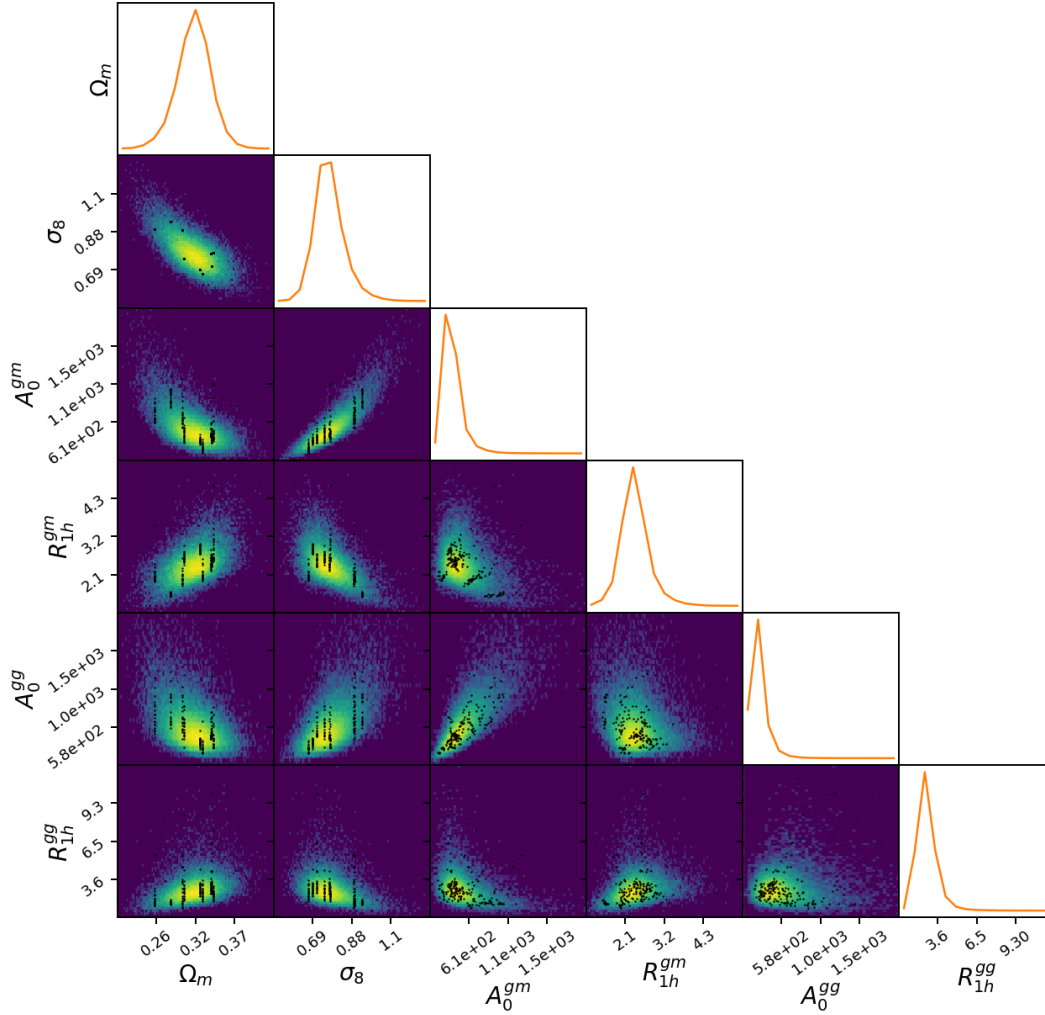
Figure 2.15: Projection of density estimated from the best-fit HZPT parameters for the LZ sample using 10 of the Aemulus simulations (Ω_m, σ_8). Best-fit parameters used for training are overplotted as black points. The R, R_1^2 and R_{2h} parameters also fitted are not shown for visibility, and are often strongly degenerate with the displayed parameters.

We also anticipate that constructing such an emulator-style tool from HZPT would require fewer training samples evaluated at distinct values of cosmological parameters. This is because the forms of the HZPT correlators are more restrictive than the form of a typical emulator.

The galaxy-galaxy and galaxy-matter A_0 parameters show a negative correlation with Ω_m and a positive correlation with σ_8, both of which are consistent with the sign of the power laws fitted for the matter power spectrum in Section 2.4 and in SV15 (for σ_8). The parameter R_{1h} seems to have the opposite relationship, which might at first appear surprising. However,

note the relationship between cosmological parameters shown in the top left panel. This is a reflection of the design strategy of the Aemulus simulations, which are based on Planck constraints [60]. Since Ω_m increases as σ_8 decreases, it is not straightforward to meaningfully disentangle dependence on Ω_m and σ_8 individually. In any event, it is clear that the HZPT parameters shown are not independent of cosmology, and depend strongly on at least some combination of Ω_m and σ_8 with scatter that is captured in the density width. We aim to provide a more quantitative comparison of the effect of different HZPT priors on cosmological constraining power in future work.

2.7   Conclusions

In this paper we expand the Halo-Zeldovich Perturbation Theory approach to accurately model two-point correlators of matter and tracers well into the nonlinear regime. For all correlators, the models are generally accurate at the percent level down to $k < 1 \; h/\text{Mpc}$ ($r > 2 \; \text{Mpc}/h$) in Fourier (configuration) space. A summary of HZPT models used in this paper is provided in Table 2.1. An additional benefit of two-point correlators in HZPT is that the corrections to the (linearly biased) ZA contributions have analytic Fourier transforms, and we provide expressions and fits for both forms of the correlators. Being analytic, this model is well suited to fast inference and gradient-based sampling, and we make the model and gradients available through a lightweight `python` package[11].

Model	Parameters	$(k_{\max}, \; r_{\min})$
$mm : n_{\max} = 2$	$\{A_0, R, R_{1h}, R_1, R_{2h}\}$	$(1, 1)$
$mm : n_{\max} = 3$	$\{A_0, R, R_{1h}, R_1, R_{2h}, R_2, R_{3h}\}$	$(8, \cdot)$
$mm : \text{alt} + n_{\max} = 2$	$\{A_0, R, R_{1h}, R_1, R_{2h}, \alpha, \beta\}$	$(10, \cdot)$
$hm : n_{\max} = 2$	$\{b_1, A_0, R, R_{1h}, R_1, R_{2h}\}$	$(1, 2)$
$hh : n_{\max} = 1$	$\{\frac{1}{\bar{n}_{\text{eff}}}, b_1, A_0, R, R_{1h}, \}$	$(1, \cdot)$
$hh : n_{\max} = 2 + \text{exc}$	$\{b_1, A_0, R, R_{1h}, R_1, R_{2h}, R_{\text{exc}}, (\sigma_{\text{exc}})\}$	$(2, 2)$
$gm : n_{\max} = 2$	$\{b_1, A_0, R, R_{1h}, R_1, R_{2h}\}$	$(1, 2)$
$gg : n_{\max} = 1$	$\{\frac{1}{\bar{n}_{\text{eff}}}, b_1, A_0, R, R_{1h}, \}$	$(1, \cdot)$
$gg : n_{\max} = 2$	$\{b_1, A_0, R, R_{1h}, R_1, R_{2h}\}$	$(\cdot, 2)$
$gg : n_{\max} = 1 + \text{exc} + 1\text{s}$	$\{b_1, A_0, R, R_{1h}, R_{\text{exc}}, \sigma_{\text{exc}}, A_{1s}, R_{1s,1h}\}$	$(\cdot, 2)$

Table 2.1: A subset of HZPT models used in this paper and their ranges of validity for different correlators. Free shot noise constants $\frac{1}{\bar{n}_{\text{eff}}}$ are only applicable for power spectra. Scales are quoted in units of ($[h/\text{Mpc}]$, $[\text{Mpc}/h]$) in the last column when available. The accuracy of all models in this table is *at least* 2%, but see individual sections for details.

---

[11]`https://pypi.org/project/gzpt/` ⚙

We demonstrate that the effect of a wide range of baryonic feedback models on the matter power spectrum - as implemented by hydrodynamical simulations - can be accounted for within the HZPT framework. These changes can be understood in terms of the halo model (see Appendix 2.8). We also provide two extended models extending the one and two-halo terms. The extended two-halo term improves upon the ZA and can reach 1% accuracy out to $k = 10\ h/\mathrm{Mpc}$ when paired with the $n_{\max} = 2$ BB term for dark matter. The extended one-halo term model is of comparable accuracy to contemporary nonlinear models (Appendix 2.8) out to $k \approx 8\ h/\mathrm{Mpc}$ for matter (including a high-feedback model of baryonic effects) and we provide power law scalings to account for variation in dark matter correlators with respect to the cosmological parameters Ω_m and σ_8.

Halo-matter and halo-halo correlators are well-described by the HZPT model when the non-perturbative phenomenon of halo exclusion is accounted for. Halo clustering is characterized in configuration space by a large-scale enhancement above $\sim 10\ h/\mathrm{Mpc}$, a non-perturbative small-scale enhancement above the exclusion scale, and a sharp step at the exclusion scale. We provide a one-parameter analytic model for the exclusion step that has analytic Fourier transform, and the small-scale enhancement is well-modeled by the BB correction term. Properly modeling halo exclusion in configuration space guarantees the correct behavior in Fourier space, including exactly accounting for sub/super-Poisson shot noise in the large-scale limit. Including shot noise in the residuals, the amplitude of the exclusion contribution with respect to the total halo-halo power spectrum is relevant at the several percent level for $k > 0.1\ h/\mathrm{Mpc}$ and the contribution is scale-dependent. We emphasize that perturbative models of halo bias that attempt to describe small scales are necessarily incomplete without a (non-perturbative) model of halo exclusion, and will fail dramatically in configuration space near the exclusion scale. We find that without explicitly modeling exclusion in the power spectrum, we obtain residuals less than 2% below $k \approx 0.7\ h/\mathrm{Mpc}$, as the $n_{\max} = 1$ BB term along with a free constant shot noise appears to account for the leading-order effect of exclusion. However, we warn that perturbative models of halo clustering that appear accurate in Fourier space at these higher k are in part modeling the scale-dependent effects of non-perturbative exclusion.

Galaxy-galaxy and galaxy-matter correlators are accurately captured by HZPT in the context of LRG-like HOD mock galaxies. Exclusion can be relevant in the simulated galaxy-galaxy correlation function for certain choices of HOD parameters, halo mass, and halo finder. We provide an estimated density that captures the relation of HZPT parameters with Ω_m and σ_8 for HOD mocks that closely resemble the BOSS LOWZ sample. Recently, emulators for tracer two-point statistics have become extremely popular [65, 139, 67, 140, 54, 141, 142, 136, 143, 144], and have provided useful and effective interpolations of simulation statistics. However, these surrogate models are usually complicated to construct and are often dependent on a number of hyperparameters, making them opaque to interpretation even beyond the inability to write down a simulation as a closed-form model. One can easily use HZPT to build a more interpretable tool similar to an emulator through the simple approach outlined in Section 2.6 using estimated priors to quantify uncertainty.

For certain galaxy samples, it is possible that exclusion may stand out in the observed

projected correlation function (depending on satellite fraction, host halo mass, or selection effects - there is perhaps a hint of this in right panel of Fig. 3 of [145]). Whether or not halo exclusion is important to include in an effective model for use in analyses is at the very least an assumption that should be checked, especially as more diverse tracers become widely used in future surveys.

One aspect of tracer auto-correlations we did not treat in this paper is the effect of cross-stochasticity, which might be especially relevant for combining populations of tracers occupying significantly different mass halos. We did not explore the redshift dependence of the model parameters for tracers, but anticipate it may be fit relatively simply as in H17. We also have not treated redshift-space distortions or other observational systematics, which of course are essential for connecting to observed two-point statistics. Finally, assuming Lagrangian density peaks are the sites of halo formation, halo exclusion will also depend on cosmology at some level [127], and so is of particular interest as measurements of cosmological information through large-scale two-point statistics of LSS are saturated and small scales remain potentially under-extracted.

HZPT serves to bring analytic descriptions of two-point correlators further into the transition and nonlinear regimes, and is a fast and interpretable complement to simulation-based models. The success of HZPT on small scales illustrates the flexibility of Padé-type expressions for modeling two-point correlators. The form of these expressions is quite simple compared to multi-loop PT, modified halo and HOD models, and most emulators. We expect that the small-scale treatment of so-called "3x2pt" analyses can be significantly improved by leveraging HZPT.

Acknowledgments

2.8 Appendix A: Details of Halo-Zeldovich Perturbation Theory

Full Model $(n_{\max} \leq 2)$

Here we provide the full HZPT model in its basic form, with explicit expressions for $n_{\max} = 0, 1, 2$.

$$P^{mm} = P_{\mathrm{Zel}} + P_{BB}^{mm} \tag{2.15}$$

$$P^{tm} = b_{tm}(P_{\mathrm{Zel}} + P_{BB}^{tm}) \tag{2.16}$$

$$P^{tt} = \frac{1}{\bar{n}_t} + b_{tt}^2(P_{\mathrm{Zel}} + P_{BB}^{tt}) \tag{2.17}$$

$$P_{BB} = A_0 \left(1 - \frac{1}{1 + k^2 R^2}\right) \frac{\sum_{n=0}^{n_{\max}-1} k^{2n} R_n^{2n}}{\sum_{n=0}^{n_{\max}} k^{2n} R_{nh}^{2n}} \tag{2.18}$$

$$n_{\max} = 0$$

$$P_{BB}(k) = F_{\mathrm{comp}}(k) = A_0 \left(1 - \frac{1}{1 + k^2 R^2}\right) \tag{2.19}$$

$$\xi_{BB}(r) = F_{\mathrm{comp}}(r) = -A_0 \frac{e^{-\frac{r}{R}}}{4\pi R^2} \tag{2.20}$$

$$n_{\max} = 1$$

$$P_{BB}(k) = P_{BB} = F_{\text{comp}}(k)\frac{1}{1 + k^2 R_{1h}^2} \tag{2.21}$$

$$\xi_{BB}(r) = F_{\text{comp}}(r)\frac{\left(1 - \left(\frac{R}{R_{1h}}\right)^2 e^{-\frac{R-R_{1h}}{RR_{1h}}r}\right)}{\left(1 - \left(\frac{R_{1h}}{R}\right)^2\right)} \tag{2.22}$$

$$n_{\max} = 2$$

$$P_{BB}(k) = F_{\text{comp}}(k)\frac{1 + k^2 R_1^2}{1 + k^2 R_{1h}^2 + k^4 R_{2h}^4} \tag{2.23}$$

$$\xi_{BB}(r) = F_{\text{comp}}(r)\left(\frac{\left(1 - \left(\frac{R_1}{R}\right)^2\right) + Ae^{r\left(\frac{1}{R} - \frac{\sqrt{R_{1h}^2 - S}}{\sqrt{2}R_{2h}^2}\right)} + Be^{r\left(\frac{1}{R} - \frac{\sqrt{R_{1h}^2 + S}}{\sqrt{2}R_{2h}^2}\right)}}{\left(1 - \left(\frac{R_{1h}}{R}\right)^2 - \left(\frac{R_{2h}}{R}\right)^4\right)}\right), \tag{2.24}$$

$$S = \sqrt{R_{1h}^4 - 4R_{2h}^4}, \tag{2.25}$$

$$A = \frac{R^2(-2R_{2h}^4 + R1^2(R_{1h}^2 - S)) + R_{2h}^4(R_{1h}^2 - S) + R_1^2(-R_{1h}^4 + 2R_{2h}^4 + R_{1h}^2 S)}{2SR_{2h}^4}, \tag{2.26}$$

$$B = -\frac{R^2(-2R_{2h}^4 + R1^2(R_{1h}^2 + S)) + R_{2h}^4(R_{1h}^2 + S) + R_1^2(-R_{1h}^4 + 2R_{2h}^4 - R_{1h}^2 S)}{2SR_{2h}^4}. \tag{2.27}$$

In this work, we reparameterize R_{2h} as $R_{2h} \equiv \frac{R_{1h}}{\sqrt{2}R_{12}}$ and vary the parameter R_{12} in our fits. We enforce $R_{12} \geq 1$, since otherwise the $n_{\max} = 2$ correlation function BB term takes on imaginary values. Analytic gradients in the python package are adjusted accordingly to be gradients of R_{12}.

We show the BB term gradients for $n_{\max} = 1, 2$ in Fig. 2.16.

Profile expansion

This closely follows Section 2 of MS14. Starting from 2.4, we expand the $j_0(kr)$ integral in its argument:

$$u_M(k) = \frac{4\pi}{M}\int_0^{R_{\text{halo}}} dr \; r^2 \; \rho_M(r)\left(1 - \frac{k^2 r^2}{3!} + \frac{k^4 r^4}{5!} - ...\right) \tag{2.28}$$

and the modulus squared of the profile is

$$|u_M(k)|^2 = |\mathcal{F}_0(M)k^0 - \mathcal{F}_1(M)k^2 + \mathcal{F}_2(M)k^4 - ...|^2 \tag{2.29}$$

where

$$\mathcal{F}_n(M) \equiv \frac{4\pi}{(2n + 1)!M}\int_0^{R_{\text{halo}}} dr \; r^{2(1+n)} \; \rho_M(r). \tag{2.30}$$
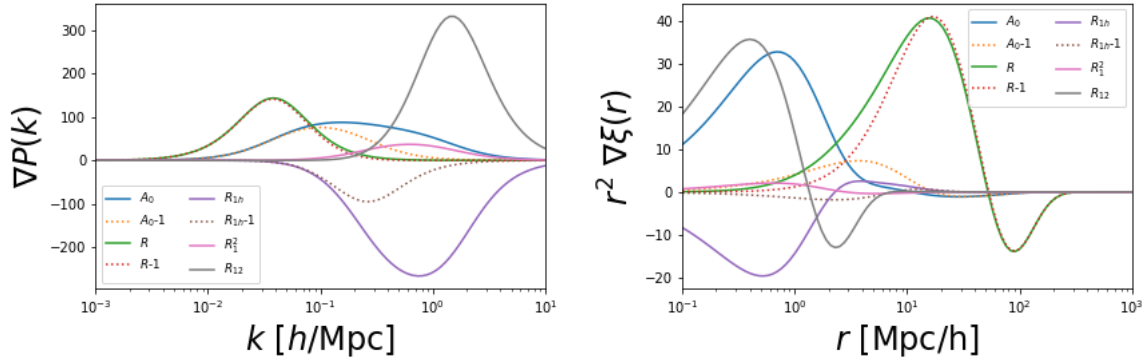
Figure 2.16: Gradients of the $P_{BB}(k)$ (*left*) and $r^2 \xi_{BB}(r)$ (*right*). Solid lines are gradients for the $n_{\max} = 2$ model, while dashed lines are for the $n_{\max} = 1$ model. For visibility for P_{BB} the A_0 and R gradients have been multiplied by 100 and 10, respectively. For visibility for ξ_{BB} the A_0 and R gradients have both been multiplied by 1000.

Then the one-halo term becomes

$$P_{1h}(k) = \int dn(M) \frac{M}{\bar{\rho}} \left(\mathcal{F}_0^2 k^0 - 2\mathcal{F}_0 \mathcal{F}_1 k^2 + (\mathcal{F}_1^2 + 2\mathcal{F}_0 \mathcal{F}_2) k^4 - ... \right), \quad (2.31)$$

which, for order $n_{\max}$, has Padé approximant with the same roots as given by the BB term in the previous section.

Introducing clarifying notation for the first few terms

$$P_{1h}(k) = A_0 (1 - \widetilde{R}_1^2 k^2 + \widetilde{R}_2^4 k^4 + ...), \quad (2.32)$$

where A_0 is the same as in the BB term. Explicitly, for $n_{\max} = 1$, the relation between the BB parameters and the profile expansion parameters is simply $R_{1h} = \widetilde{R}_1$, while for $n_{\max} = 2$ the relations are the following:

$$R_{1h}^2 = \left(\frac{\widetilde{R}_1^2 \widetilde{R}_2^4}{\widetilde{R}_1^4 - \widetilde{R}_2^4} \right), \quad (2.33)$$

$$R_{2h}^4 = \left(\frac{\widetilde{R}_2^8}{\widetilde{R}_1^4 - \widetilde{R}_2^4} \right) \quad (2.34)$$

,

$$R_1^2 = \left(\frac{2\widetilde{R}_1^2 \widetilde{R}_2^4 - \widetilde{R}_1^6}{\widetilde{R}_1^4 - \widetilde{R}_2^4} \right). \quad (2.35)$$

For $R_1^2 \geq 0$, we have $\widetilde{R}_1 \geq \widetilde{R}_2 \geq \frac{\widetilde{R}_1}{\sqrt{2}}$

Deriving $R_{n,nh}$ from the halo model with baryonic effects

To illustrate that it is possible to derive the HZPT $R_{n,nh}$ profile parameters in the context of baryonic effects from the halo model, we calculate them for the simple case of the mass function of [59] (M200c) and NFW profile [100] (with concentration $c = 5$ for simplicity). We approximate the impact of baryons on the HZPT parameters via "baryonified" profiles as modeled in [152] (and constrained by X-ray data). These profiles incorporate the presence of stellar mass and satellite galaxies, gas that has been pushed by feedback toward the edge of the halo, and the resulting response of the dark-matter profile to these changes (see Section 2 of [152]). We fix the free parameters of the baryon-matter profile to the best-fit values of [152] corresponding to the "Model B-avrg" scenario: $\eta_{cga} = 0.6$, $\eta_{star} = 0.32$ for the stellar profile, and $M_c = 6.6 \times 10^{13} M_\odot/h$, $\mu = 0.21$, $\theta_{ej} = 4$ in the determination of the gas profile.

Performing the integrals of eqn. 2.30, we find, for the dark matter only (dmo) profile, that $(A_0, R_{1h}, R_1^2, R_{2h})_{\mathrm{dmo}} = (1114\ [h/\mathrm{Mpc}]^3,\ 6.1\ \mathrm{Mpc}/h,\ 35\ [\mathrm{Mpc}/h]^2,\ 2.9\ \mathrm{Mpc}/h)$, and, for the baryon and dark matter profile (dmb), that $(A_0, R_{1h}, R_1^2, R_{2h})_{\mathrm{bdm}} = (1112\ [h/\mathrm{Mpc}]^3,\ 5.1\ \mathrm{Mpc}/h,\ 24\ [\mathrm{Mpc}/h]^2,\ 2.6\ \mathrm{Mpc}/h)$.

The difference in A_0 is negligible, as we would expect since A_0 is essentially the one-halo amplitude (which was fixed in Section 2.4 along with R). However, the relative changes in the $R_{n,nh}$ parameters are significant, and are similar to those discussed in Section 2.4 (though the correspondence is not exact given the idealized setting). Here we may clearly identify the source of the change in parameters (after translating mass-integrated profile moments $\tilde{R}_n$ to Padé parameters $R_{n,nh}$). The $\mathcal{F}_1(M)$ and $\mathcal{F}_2(M)$ values both grow uniformly in mass with respect to the dmo case, since in the dmb case the gas is pushed out toward the outskirts of the halo, where matter contributes more to the r^4, r^6 moments.

The radius at which to truncate the $\mathcal{F}_n$ integrals is somewhat unclear, but here we use a scale close to the truncation radius of the NFW profile for definiteness (with $R_{\mathrm{halo}} = \frac{9}{8}\epsilon r_{200c}$ with $\epsilon = 4$). This choice of truncation radius results in good agreement in the enclosed masses of the dmo and dmb profiles, and is small enough to prevent the integrals from being sensitive to two-halo contributions (as modeled in [152]). There is a very small difference ($\sim 0.1\%$) in enclosed masses that is the source of the negligible difference in A_0 for the two profiles. Given the uncertain nature of realistic matter profiles (e.g. scatter in the concentration-mass relation), it remains advantageous to take an agnostic attitude toward halo profile details and avoid integrating directly as we do in the main text.

Comparison to `HMCode2020`

Figure 2.17 shows residuals of matter power spectrum predictions from `HMCode2020` [119] with respect to the test set drawn from the Mira-Titan CosmicEmu [66]. The `HMCode2020` predictions are as provided through `CAMB` [150]. Compared with Fig. 2.6, the residuals are similar on the largest scales, slightly larger residuals on quasi-linear scales, and very slightly smaller on the smallest scales. The residuals on quasi-linear scales appear larger than presented in Fig. 2 of [119], but are consistent with residuals with respect to Mira-

Titan quoted in their Fig. D1 (we have CosmicEmu in the numerator, so our residuals are inverted with respect to theirs). The largest residuals at low k trend with low-σ_8 models and the largest residuals at 0.1 $h/\mathrm{Mpc} < k < 1$ h/Mpc trend with both low-σ_8 models and high-Ω_{cb} models. From this comparison, it is clear that the HZPT model is competitive when varying σ_8 and Ω_{cb}.
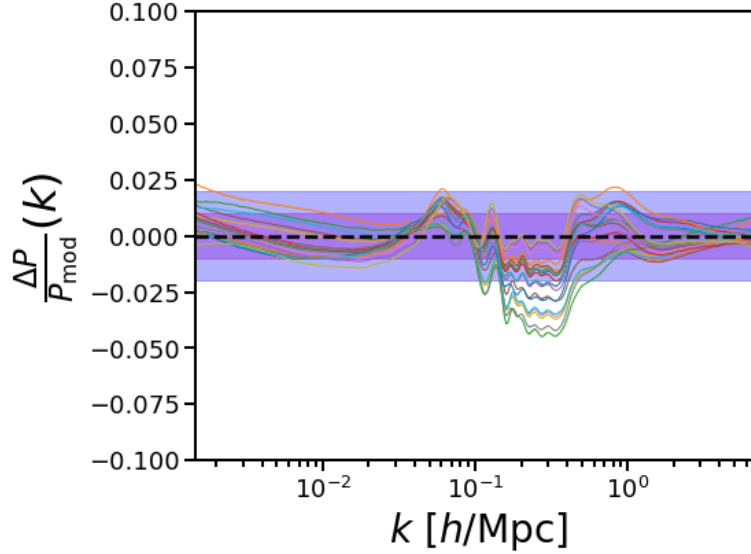
Figure 2.17: Residuals for `HMCode2020` with respect to the same CosmicEmu test set shown in the right panel of Figure 2.6.

2.9   Appendix B: Halo correlator model components

Model for ξ_{hh} and P_{hh} including halo exclusion

Considering a general configuration-space model for the exclusion step $F_{\mathrm{exc}}(r)$, the two-point function of the discrete halo field is modeled as

$$\xi_{hh}^d(r) - \frac{1}{\bar{n}}\delta^D(\mathbf{r}) = F_{\mathrm{exc}}(r)\left[1 + \xi_{hh}^c(r)\right] - 1 \tag{2.36}$$

which gives in Fourier space

$$P_{hh}^d(k) = \frac{1}{\bar{n}} + \mathcal{F}\left(F_{\mathrm{exc}}(r)\left[1 + \xi_{hh}^c(r)\right] - 1\right) \tag{2.37}$$

$$= \frac{1}{\bar{n}} + \left[\mathcal{F}\left(F_{\mathrm{exc}}(r)\right) * \mathcal{F}\left(1 + \xi_{hh}^c(r)\right)\right](k) - \delta^{(D)}(k) \tag{2.38}$$

$$= \frac{1}{\bar{n}} + F_{\mathrm{exc}}(k) + \left[F_{\mathrm{exc}}(k) * P^{(c)}(k)\right], \tag{2.39}$$

where we have dropped the zero-lag Dirac delta in the final expression.

We may also write this expression in a slightly clearer way conceptually, conforming to the convention of B13 (see eqn. 2.36). Instead of specializing to the too-simple top-hat model for the exclusion step (c.f. Figure 2.9), we assume only that we may write $F_{\text{exc}}(r) = 1 - W(r)$, where $W(r)$ is a function with analytic Fourier transform. Then, the expression is identical to eqn. 2.12, but we have taken the tiny step of generalizing the top-hat window to something more realistic:

$$P_{hh}^d(k) = \frac{1}{\bar{n}} + P^{(c)}(k) - \tilde{W}(k) - \left[\tilde{W} * P^{(c)} \right](k). \tag{2.40}$$

Also, we may define a finite-size stochastic contribution as the deviation of the true discrete power spectrum from the continuous model and the fiducial (Poisson) constant arising from discreteness:

$$S(k) = P^{(d)}(k) - \frac{1}{\bar{n}} - P^{(c)}(k) \tag{2.41}$$

$$= -\tilde{W}(k) - \left[\tilde{W} * P^{(c)} \right](k) \tag{2.42}$$

This is the quantity plotted in the right-hand panel of Figure 2.9, which quantifies the failure to include a model for exclusion in Fourier space.

As described in Section 2.5, we consider two effective models for the halo exclusion step, both of which perform much better than the simple top-hat. The ErfLog model (almost the same as B13) is:

$$F_{\text{exc}}(r) = \frac{1}{2} \left[\text{erf} \left(\frac{\log_{10} \left(\frac{r}{R_{\text{exc}}} \right)}{\sigma_{\text{exc}}} \right) + 1 \right] \tag{2.43}$$

and the Exp model is

$$F_{\text{exc}}(r, R_{\text{exc}}) = \left[1 - \exp \left(- \left(\frac{r}{R_{\text{exc}}} \right)^4 \right) \right]^2. \tag{2.44}$$

The ErfLog model has no analytic Fourier transform, which must be computed numerically. However, the Exp model has analytic Fourier transform, where the Fourier transform of the squared quantity is given by (defining the function f):

$$\delta^{(D)}(k) + f(k, R_{\text{exc}}) = \delta^{(D)}(k) - \frac{R_{\text{exc}}^3}{3} \Gamma\left(\frac{7}{4}\right) {}_0F_2\left(\frac{1}{2}, \frac{5}{4}; \left(\frac{kR_{\text{exc}}}{4}\right)^4\right)$$
$$+ \frac{k^2 R_{\text{exc}}^5}{24} \Gamma\left(\frac{5}{4}\right) {}_0F_2\left(\frac{3}{2}, \frac{7}{4}; \left(\frac{kR_{\text{exc}}}{4}\right)^4\right), \quad (2.45)$$

where ${}_0F_2$ is the generalized hypergeometric function. The full expression is then $F_{\text{exc}}(k) = \delta^{(D)}(k) - 2f(k, R_{\text{exc}}) + f(k, 2^{-\frac{1}{4}} R_{\text{exc}})$. It is helpful to consider this function from the perspective of modeling the nonperturbative exclusion effect in Fourier space (i.e. in the context of a

Fourier space analysis), and in clearly disentangling the shape of the terms due to finite halo size and due to nonlinear clustering. This form of the model is potentially computationally inefficient, but is presented for conceptual completeness, and convolutions may be sped up via FFTLog-based algorithms (e.g. [153]).

Halo mass bins

The halo bins used here are provided in Table 2.2. We do not show bins 0 and 3 in the main

Bin	$[\log M_{\min}, \log M_{\max}]$
1	[12.5, 13.5]
2	[13.5, 14.5]
3	[11.5, 12.0]
4	[12.0, 12.5]
5	[12.5, 13.0]
6	[13.0, 13.5]
7	[13.5, 14.0]
8	[14.0, 14.5]

Table 2.2: Halo mass bins in $M_\odot/h$

text since these low mass halos are almost certainly very poorly-resolved. We also show the halo-matter and halo-halo correlation functions for wider bins that are the combination of the narrower two bins described in the main text (Fig. 2.18, Fig. 2.19). The width of the bins seems not to adversely affect the accuracy of the model.

Compensation R for halos

Fixing R for halo auto-correlation to some very large value (10^{10}, essentially infinity for our purposes) has no real effect (fraction of 1%) on the quality of fit in the halo auto-correlation for the $n_{\max} = 1$ model with free shot noise. For halo-matter, the scale at which the compensation is relevant is different for different sized halos and increases with halo mass (we find this in the fitted values of R for halo-matter). Fixing R to any particular value that works well for a certain mass bin produces significantly worse fits on large scales for other mass bins - so the compensation is not something that can easily be fixed for the cross correlation. Disentangling compensation from nonlinear bias is also challenging since we expect both the physical compensation scale and nonlinear bias to change with halo mass.
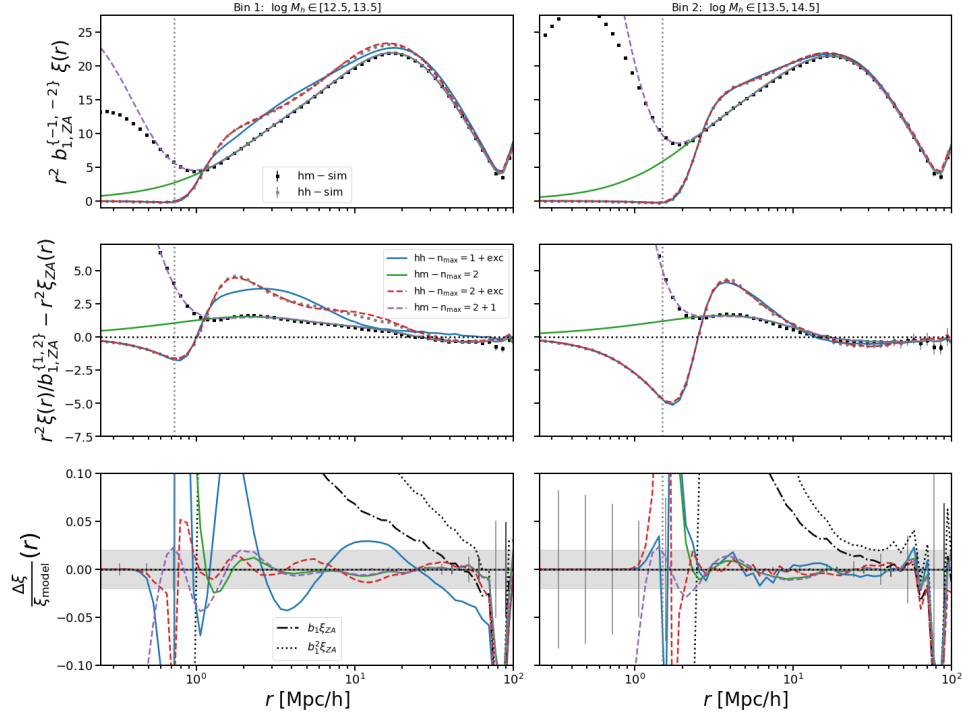
Figure 2.18: Same as Fig. 2.7, but for the combined bins.
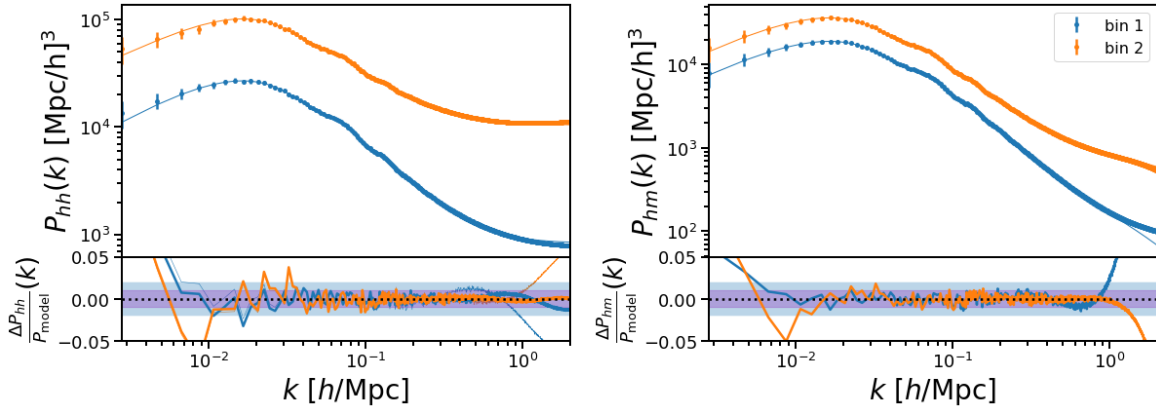
Figure 2.19: Same as Fig. 2.8, but for the combined bins.

2.10   Appendix C: Impact of Halo Finder and Mass Definition on Halo Exclusion

The halo finder used has a significant impact on the scale and amplitude of the exclusion feature in the halo-halo correlation function. The halo mass definition also has a moderate

effect on the scale and significant impact on the amplitude of the feature.

We saw in Section 2.6 that the exclusion feature presented itself at much smaller scales (≈ 0.7 Mpc/h) in the LZ correlation functions than in the CM correlation functions (≈ 1.5 Mpc/h). To check if this effect is due to the choice of halo finder, we ran FoF on the Aemulus dark matter particle snapshots that were used to create the ROCKSTAR strict SO M200b halos provided in the Aemulus halo catalogs. Fig. 2.20 shows the correlation function for both FoF halos and ROCKSTAR strict SO (SSO) halos. It is clear that the correlation function of FoF halos exhibits an exclusion feature at much larger scales that the one for SSO halos does. To check that this is not an effect unique to the Aemulus simulations, and
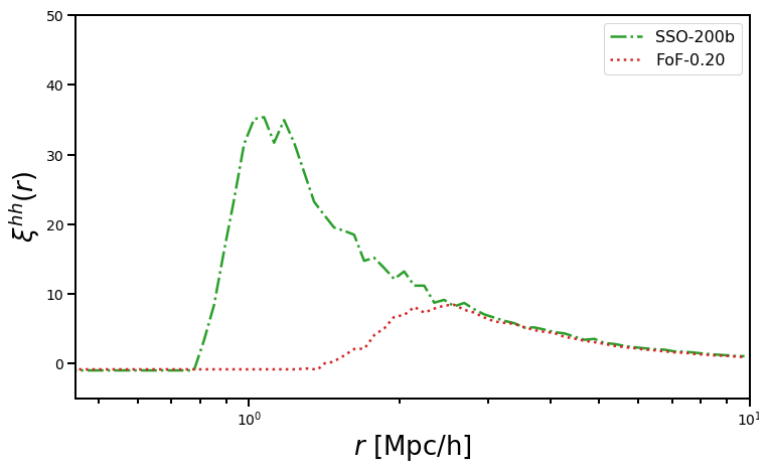
Figure 2.20: Halo-halo correlation function for the default Aemulus ROCKSTAR "strict SO" (SSO) M200b halos (includes unbound particles) in green, and for FoF (linking length 0.2) halos in red.

to provide a direct comparison of SO and SSO mass definitions using the same halo centers, we use the publicly available Abacus halo catalogs [56]. The halo catalogs used correspond to Planck cosmology with box size 1100 Mpc/h at $z = 0.5$, and details of the FoF and ROCKSTAR catalogs are provided in [56]. Fig. 2.21 shows the same as Fig. 2.20 for the Abacus halos, but with additional curves for SO virial halos as well as default ROCKSTAR virial halos. We consider only bin 7 here, but this effect persists for all halo mass bins considered in Table 2.2 (the only real differences being the absolute scales and smoother curves for lower-mass halos due to increased number density at lower mass).

It appears that for strict SO halos, halo mass definition does not change the exclusion feature significantly. However, the use of default ROCKSTAR halo masses instead of SSO masses is already quite different in terms of the exclusion feature. The exclusion features for the ROCKSTAR halo masses lie in the middle of those of the SSO and FoF halos. The change in amplitude along with the change in scale is perhaps not that surprising, as it qualitatively seems to be what would be expected of applying the exclusion step to the continuous model

for $\xi_{hh}(r)$ described in Section 2.5 at a smaller R_{exc}. Naively this would seem to mean that the SSO halos display clustering behavior indicative of a smaller exclusion scale (at fixed halo finder) than that exhibited by the ROCKSTAR halos (and the same goes for FoF halos). This may be related to the well-known scatter in the relation between FoF ($b = 0.2$) and M200b halos (e.g. [154],[155]). It would be interesting to connect this difference with a more intuitive description of exclusion, perhaps informed by peaks similar to [128]. Figs.
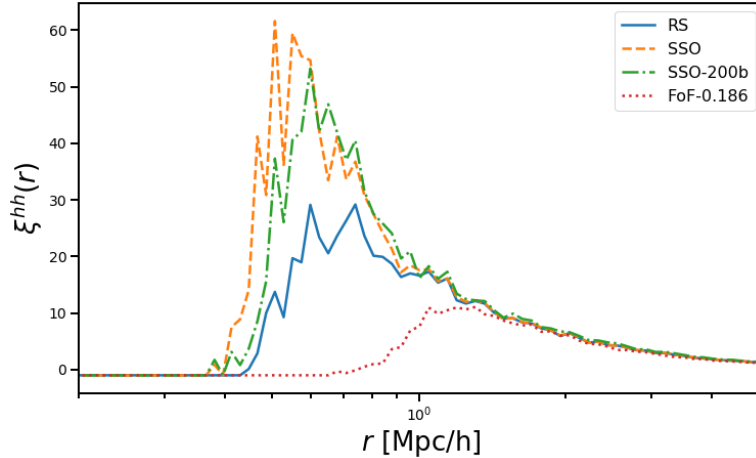
Figure 2.21: Halo-halo correlation function for the Abacus ROCKSTAR default halos (blue), "strict SO" (SSO) virial halos (orange), SSO M200b halos (green), and for FoF with linking length 0.186 halos (red).

2.20 and 2.21 show features that are similar to Fig. A1 of [156] that compares different percolation strategies in a modified version of ROCKSTAR (but did not investigate $\xi_{hh}(r)$ in detail). It would be interesting to draw clearer connections between the differences in percolation strategies, the differences we observe here, and the effect on halo exclusion in two-point correlators.

2.11 Appendix D: Cross-correlation coefficient r_{cc}

The cross correlation coefficient provides an additional view of the non-perturbative effects present in small-scale two point statistics. The cross correlation coefficient for two-point statistics is defined as $r_{cc}^{xm}(X) = \frac{X^{xm}}{\sqrt{X^{mm}X^{xx}}}$, where x denotes the tracer type (galaxies or halos) and X is the two-point statistic in question P or ξ. We present cross-correlation coefficients measured from the simulations and from the HZPT model for halos and galaxies in Figure 2.22. The cross correlation coefficient for halos and galaxies is very close to 1 for all halo mass bins and HOD parameter choices in configuration space above $r \sim 4$ Mpc/h. On smaller scales, non-perturbative effects become obvious - the cross-correlation coefficient

first drops in the region of the SSE and attains a minimum before blowing up as r drops below r_{exc} at the exclusion step before becoming undefined as $\xi^{hh}(r)$ crosses zero.

For HOD galaxies, the picture is similar to the case of halos in a mass bin around $\log \frac{M}{M_\odot/h} \in [13, 13.5]$, but with significant scatter due to the details of the halo occupation. For many choices of halo occupation, this leads to $r_{cc}^{gm}(r) > 1$ (but with a less severe growth toward small scales than for halos). The undefined behavior at small scales for halos is also no longer present for galaxies since the satellite profile gives a non-zero contribution to $\xi_{gg}(r)$ on the scales at which $\xi_{hh}(r) \to -1$.

In Fourier space, the situation is perhaps more complicated due to finite halo size effects that manifest partially as sub/super-Poisson shot noise - a more complete picture of finite-halo size effects is given by the modeling described in Section 2.5. By $k \sim 0.1 \ h/\mathrm{Mpc} \ r_{cc}(k)$ already deviates from unity significantly for both halos and galaxies. For halos, there is an offset from unity due to finite halo size ("sub/super-Poisson shot noise" on large scales) that is more pronounced for larger halos, as expected. Higher mass halos give $r_{cc}^{hm}(k)$ with a weaker scale dependence, leveling off quickly compared to lower mass halos, which continue to drop as k increases.
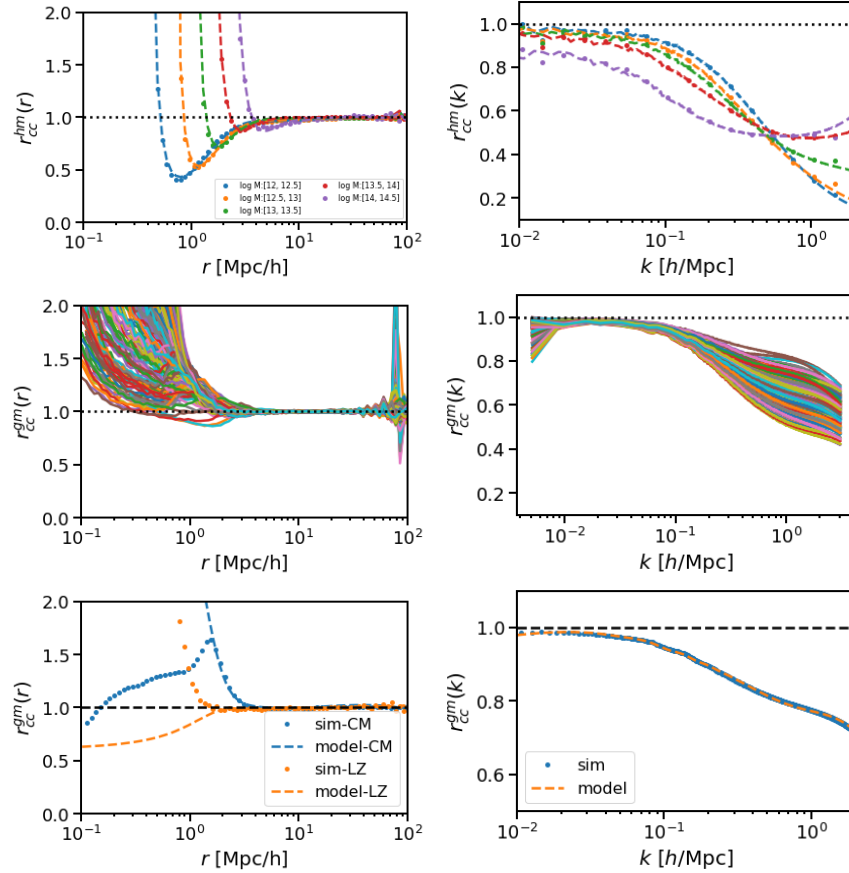
Figure 2.22: *Top Left:* Cross-correlation coefficient for halos in configuration space. *Top Right:* Cross-correlation coefficient for halos in Fourier space. *Middle Left:* Cross-correlation coefficient for (LOWZ) HOD galaxies in configuration space. *Middle Right:* Cross-correlation coefficient for (LOWZ) HOD galaxies in Fourier space. Shot noise is *not* subtracted for the Fourier space cross-correlation coefficients. *Bottom Left:* Cross-correlation coefficient and model for HOD galaxies in configuration space for single LOWZ and CMASS HOD realizations. *Bottom Right:* Cross-correlation coefficient and model for HOD galaxies in configuration space for single CMASS HOD realization. Shot noise is *not* subtracted for the Fourier space cross-correlation coefficients.

Chapter 3

Accurate and high-performance numerical simulations with neutrinos

In this chapter, we update existing methodology for simulating the impact of massive neutrinos on the cosmological density field using high-precision numerical methods. Neutrino mass constraints are a primary focus of current and future large-scale structure (LSS) surveys. Non-linear LSS models rely heavily on cosmological simulations – the impact of massive neutrinos should therefore be included in these simulations in a realistic, computationally tractable, and controlled manner. A recent proposal to reduce the related computational cost employs a symmetric neutrino momentum sampling strategy in the initial conditions. We implement a modified version of this strategy into the Hardware/Hybrid Accelerated Cosmology Code (HACC) and perform convergence tests on its internal parameters. We illustrate that this method can impart $\mathcal{O}(1\%)$ numerical artifacts on the total matter field on small scales, similar to previous findings, and present a method to remove these artifacts using Fourier-space filtering of the neutrino density field. Moreover, we show that the converged neutrino power spectrum does not follow linear theory predictions on relatively large scales at early times at the 15% level, prompting a more careful study of systematics in particle-based neutrino simulations. We also present an improved method for backscaling linear transfer functions for initial conditions in massive neutrino cosmologies that is based on achieving the same relative neutrino growth as computed with Boltzmann solvers. Our self-consistent backscaling method yields sub-percent accuracy in the total matter growth function. Comparisons for the non-linear power spectrum with the Mira-Titan emulator at a neutrino mass of $m_\nu = 0.15$ eV are in very good agreement with the expected level of errors in the emulator and in the direct N-body simulation.

3.1 Introduction

The non-zero mass of neutrinos is perhaps the most well-established manifestation of beyond Standard Model physics [157, 158, 159]. Constraints on total neutrino mass are available from terrestrial experiments with a current upper limit of 0.8 eV, which is expected to be eventually reduced to a mass sensitivity of ~ 0.2 eV [160]. Current cosmological limits from the cosmic microwave background and large-scale structure (LSS) surveys have tighter bounds, at the level of 0.1 eV [3, 161] with some sensitivity to the modeling of systematic effects. LSS observations play an important role in constraining neutrino masses due to the characteristic suppression of the matter power spectrum, $P_m(k)$, below the neutrino free-streaming scale. This scale is roughly $k_{fs}(z) \approx (\frac{m_\nu}{\text{eV}})\sqrt{\Omega_m a}\ h\ \text{Mpc}^{-1}$ in matter domination [2]. Since the free-streaming scale is proportional to the neutrino mass, m_ν, and the total amplitude of suppression is proportional to the neutrino energy density, Ω_ν, observational probes of large-scale structure afford the opportunity, at least in principle, to probe both the individual and total neutrino masses.

The approach to constraining neutrino mass based on cosmological observations complements particle physics-based probes such as oscillation experiments [e.g., 162, 163] that are sensitive to mass-squared differences between mass eigenstates and tritium β-decay experiments [164, 165, 166, 167] that aim to constrain the absolute neutrino mass scale. Although disentangling individual neutrino masses from cosmological observations remains challenging in practice [168], upper bounds placed on the total mass are becoming increasingly precise. For instance, DESI is expected to measure the total mass with an error of 0.02 eV [169] or less when combined with CMB-S4 [170], which is an order of magnitude improvement over the lower limit of Ref. [160].

The uncertainty on measurements of matter two-point statistics in modern surveys is small enough that the effect of massive neutrinos needs to be included for accurate modeling thereof as measured, e.g., by gravitational lensing or spectroscopic galaxy surveys [171, 172, 173, 38, 174, 175, 39, 176, 177, 178, 179]. The effect of neutrinos on LSS has also been highlighted through alternative probes such as voids [180, 181, 182, 183, 184], which have been shown to be particularly sensitive to neutrinos in simulations (however, see Ref. [185]), or through higher-order statistics like the bispectrum [186, 187, 188]. Accurate modeling of the total matter density field over a wide range of scales is required to exploit these probes. In particular, it is well-known that neutrinos produce scale-dependent growth of the matter field on scales below their free-streaming scale, which is of order $k_{\text{fs}} \sim 0.1\ h\ \text{Mpc}^{-1}$ for a neutrino of mass $m_\nu \sim 0.1$ eV (at $z = 0$). However, the detailed effect of massive neutrinos on the small-scale matter distribution remains less well understood due to non-linear evolution at $k \gtrsim 0.1\ h\ \text{Mpc}^{-1}$, especially for statistics beyond the matter power spectrum. While halo model based approaches can be pursued [189, 190], more accurate modeling of the matter field on these nonlinear scales is attainable, in principle, through N-body simulations that model the growth of structure in the presence of massive neutrinos.

Cosmological simulations may account for the presence of massive neutrinos at several

levels of detail[1]. The simplest prescription is to treat massive neutrinos explicitly only at
the level of the homogeneous and isotropic background, and including the neutrino per-
turbations through linear theory, effectively a leading-order expansion in the parameter
$f_\nu = \Omega_\nu/\Omega_m$ [191, 192, 193]. This method has been used in the Mira-Titan power spec-
trum emulator [194], to which we will compare our results in Section 3.4. At the next level,
massive neutrino perturbations are evolved on a mesh following linear theory and are added
as a background component to the gravitational potential during the cold matter (cold dark
matter and baryons) force calculation [195]. This approach is potentially problematic, as the
choice to evolve cold matter non-linearly while evolving massive neutrinos at the linear level
violates momentum conservation [196], in addition to obviously discarding the impacts of
non-linear neutrino growth. Various classes of "linear response" methods [197, 198, 199] take
this another step further by allowing the background neutrino mesh to evolve perturbatively
in response to the instantaneous non-linear cold matter field of the simulation. Alternative
simulation techniques have also been proposed including the computationally ambitious ap-
proach of directly solving the Vlasov-Poisson equation in phase space [200], or to solve the
neutrino evolution with a Boltzmann hierarchy expansion [201, 202, 203].

Perhaps the most accurate, yet still feasible, choice for non-linear LSS modeling is to
treat both cold species and massive neutrinos on equal footing in multi-species particle
simulations. The idea of multi-species simulations, with separate particles for the cold matter
and neutrinos, is far from new [e.g., 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214,
215, 216] and, while superficially straightforward, has a number of subtleties leading to
variations in its implementation. A comparison campaign between a number of particle-based
approaches as well as some approximate techniques has been recently performed in Ref. [217].
Variations in the particle method typically stem from numerical difficulties associated with
modeling fast-moving thermal neutrino particles at early times. One way to overcome this
challenge is to adopt a "hybrid" approach [218, 219, 220] where the mesh-based strategies
described above are utilized at high redshift while switching to particle-based schemes at
later times when the neutrino thermal motion has been sufficiently damped. Given the
motivation to constrain neutrino masses via cosmological observations, there has also been
a recent effort to extend the physical fidelity of neutrino simulations by including baryons
as an additional species treated numerically with hydrodynamics [221, 222, 223].

One of the main challenges posed by neutrino particle simulations is that they are typi-
cally limited by shot noise effects due to the high thermal motion of neutrinos at early times.
These effects overwhelm the cosmological neutrino signal in the power spectrum unless an
extreme number of neutrino particles are simulated [212]. To address this issue, Ref. [224]
introduced a method of setting initial conditions for neutrino particles that involves a sym-
metric sampling of the neutrino momenta across the simulation domain. This symmetry
drastically reduces the scale-independent shot noise that has plagued earlier simulations,
allowing the use of many fewer neutrino particle tracers. For this reason, a variant of this

---

[1]The following techniques can be applied to any massive thermal relic species, but we will restrict our
attention to massive neutrinos.

symmetric scheme (hereafter "tiling", to refer to the angular tiling of momenta on the sphere) has been explored by other particle-based neutrino simulations (e.g., Refs. [213, 214, 215]). However, while tiling succeeds in suppressing neutrino shot noise, it can also introduce numerical artifacts in *both* the neutrino and cold matter density fields, as well as their growth, as we will show.

Another challenge with neutrino simulations relates to how the Newtonian approximation complicates the construction of initial conditions. To initialize the cold matter and neutrinos, linear theory transfer functions, generally computed using a Boltzmann solver, must be provided to the simulation as input. To recover the correct observationally-relevant linear matter power spectrum on large scales at low redshift, one cannot simply provide the transfer functions at the initial redshift of the simulation. Boltzmann codes (correctly) include non-negligible radiation and metric perturbation terms in the linear equations of motion of the evolved species, which are not included in the Newtonian simulation. It is possible to include the radiation (and massive neutrino) perturbations explicitly when performing the N-body simulation [225, 226], but usually some form of "backscaling" is employed instead [227]. For simulations including massive neutrino particles, a linear two-fluid approximation for backscaling transfer functions to account for scale-dependent growth of matter due to neutrinos was developed in Ref. [228]. Massive neutrinos, however, do not constitute a fluid. We address the benefits and shortcomings of treating them as such, and introduce our own improved backscaling scheme.

In this paper, we focus on the two main challenges outlined above. First, we present an iterative backscaling method in Section 3.2 that is designed to create a self-consistent initial condition framework for massive neutrino cosmologies. Next, we follow the tiling approach [224], by implementing it within the Hardware/Hybrid Accelerated Cosmology Code (HACC; [229]). We describe our custom implementation of the tiling scheme in Section 3.3 along with the extra code modifications required to efficiently simulate neutrino particles in a high-performance setting. We present a suite of neutrino simulations in Section 3.4 that are used to perform a numerical convergence study on the internal parameters of the tiling scheme. We also show results from a mitigation strategy used to remove numerical artifacts arising from the discretization of the initial condition neutrino grid. We compare our final results for the total matter power spectrum with the Mira-Titan emulator and find good agreement. Furthermore, we use our backscaled neutrino transfer functions to show that the simulated neutrino power spectrum systematically deviates from linear theory at early times. We finish with concluding remarks in Section 4.5.

3.2 Iterative Backscaling

Setting up the initial conditions of the cold matter and neutrino particles requires the density and velocity fields at the starting redshift of the simulation, z_{ini}. These fields can be obtained in linear theory using a Boltzmann solver such as CAMB [230] or CLASS [231]. The only caveat is that cosmological simulations typically do not perform the correct linear evolution between

the time the species are initialized and the final simulation redshift, $z_{\rm fin}$. In particular, simulations generally employ a Newtonian forward model that omits the radiation and metric perturbations included in the evolution equations of Boltzmann codes. As such, the usual convention is to construct the initial conditions in such a way that the final growth in density perturbations matches the linear theory total matter power spectrum, $P_m(z_{\rm fin}, k)$, at the end of the simulation. This process is referred to as "backscaling". For $z > z_{\rm fin}$, the simulation and Boltzmann code power spectra will disagree at the linear level, with this disagreement greatest on large scales and early times (see Figure 3.2 for an example). In the absence of massive neutrinos, the growth of total matter density perturbations in the Newtonian forward model will be scale-independent meaning that the $z_{\rm fin}$ linear transfer function can simply be rescaled using a multiplicative factor, D, known as the growth factor.

The procedure for computing the scale-independent growth factor in a massless neutrino cosmology involves solving the coupled set of continuity, Euler, and Poisson equations, expressed below at the linear level [e.g., 232]:

$$\frac{\partial \delta_m}{\partial a} + \frac{\theta_m}{a^2 H} = 0, \tag{3.1}$$

$$\frac{\partial \theta_m}{\partial a} + \frac{\theta_m}{a} = -\frac{1}{a^2 H} \nabla^2 \phi, \tag{3.2}$$

$$\nabla^2 \phi = \frac{3}{2} H^2 \Omega_m a^2 \delta_m, \tag{3.3}$$

where $\theta \equiv \nabla \cdot \mathbf{v}_m$ is the velocity divergence field and the subscript m denotes the combined matter field (cold dark matter plus baryons). The solution[2] to these equations is found by assuming that δ_m can be factored into a component that depends only on space and a component that depends only on time. The time-dependent component is known as the growth factor since it describes the growth of density perturbations in time: $\delta_m(a) \propto D(a)$. It follows from the continuity equation that $\theta_m(a) \propto a^2 H(a) \mathrm{d}D/\mathrm{d}a$. From the above expressions, it is clear that the growth factor depends on the particular cosmology, via the background expansion in the Hubble parameter H, as well as through the gravitational source term of the Poisson equation, set by the total matter density, Ω_m. For reasons made clear below, we emphasize the latter dependence by writing the growth factor as $D(a, \Omega_m)$.

Once the growth factor has been computed, backscaling is accomplished by rescaling the transfer function that encodes the time evolution of matter perturbations in linear theory obtained from the Boltzmann code at the final redshift, $T_m(z_{\rm fin}, k)$, in the following manner:

$$T_m^{\rm bs}(z_{\rm ini}, k) = \frac{D(z_{\rm ini}, \Omega_m)}{D(z_{\rm fin}, \Omega_m)} T_m(z_{\rm fin}, k), \tag{3.4}$$

where the superscript "bs" denotes the backscaled transfer function used as an input to the simulation initial conditions. In the event that cold dark matter and baryon particles

---

[2]There are two independent solutions to this set of equations but we focus only on the "growing" mode while ignoring the "decaying" mode that becomes subdominant at late times.

are treated as separate species (e.g., in hydrodynamics simulations) then their respective backscaled transfer functions follow as:

$$T_\alpha^{\mathrm{bs}}(z_{\mathrm{ini}}, k) = F_\alpha(z_{\mathrm{ini}}, k) \frac{D(z_{\mathrm{ini}}, \Omega_m)}{D(z_{\mathrm{fin}}, \Omega_m)} T_m(z_{\mathrm{fin}}, k), \tag{3.5}$$

where the subscript $\alpha \in \{c, b\}$ denotes cold dark matter and baryons, respectively, and we define the term

$$F_\alpha(z, k) \equiv T_\alpha(z, k)/T_m(z, k). \tag{3.6}$$

It is important that F_α is used in this manner to ensure that each species is initialized with the correct proportion relative to the total matter field at z_{ini} (i.e., backscaling by simply replacing T_m in equation (3.4) with T_α would be incorrect due to the fact that the individual solutions for $\delta_{c,b}$ cannot be factored into spatial- and time-independent components as could be done with δ_m). The initial particle velocities are computed by defining an initial velocity transfer function, $T_{\theta,\alpha}$, for each species. This is computed from the continuity equation as the time derivative of the backscaled density transfer function evaluated at the initial time:

$$T_{\theta,\alpha}^{\mathrm{bs}}(z_{\mathrm{ini}}, k) = -a^2 H(a) \left. \frac{\mathrm{d}T_\alpha^{\mathrm{bs}}}{\mathrm{d}a} \right|_{z=z_{\mathrm{ini}}}. \tag{3.7}$$

In practice, this can be computed using a finite difference of equation (3.4) evaluated at $z_{\mathrm{ini}} \pm \epsilon$ [206, 211] (here we use $\epsilon = 0.1$). If only the combined matter field is desired, as is the case in standard single-species simulations, then it follows from equation (3.4) that the time derivative of T_m^{bs} in equation (3.7) reduces to the time derivative of D, meaning that the velocity transfer function is proportional to the density transfer function. For this reason, single-species simulations generally do not use an explicit velocity transfer function, but rather just numerically evaluate the time derivative of D.

In the presence of massive neutrinos, backscaling is complicated by the fact that the total matter field no longer evolves in a scale-independent manner, even in the Newtonian forward model. The reason is related to the large thermal component of neutrino motion that induces a characteristic free-streaming scale, k_{fs}. In short, neutrinos cluster like cold matter on scales $k \ll k_{\mathrm{fs}}$ but possess sufficient kinetic energy to evade capture in the gravitational potential sourced by cold matter on scales $k \gg k_{\mathrm{fs}}$. In other words, neutrinos contribute to the growth of density perturbations only on scales much larger than the free-streaming scale. Hence, the usual backscaling prescription presented above would only be valid on the large scales for which neutrinos contribute to the growth of density perturbations.

In principle, the solution to this problem is to solve the coupled set of growth equations for both the cold matter and neutrinos in order to compute the scale-dependent growth factor that can be used in equation (3.4). The difficulty with this approach is that the neutrino Euler equation includes a pressure gradient term that is not trivially expressible in an analytic framework. One strategy is to model neutrinos using the fluid approximation so that a tractable solution can be obtained. This is the approach taken in the code REPS [228] which has attained widespread usage in cosmological simulations [214, 233, 215, 221, 234,

185, 235, 236, 144, 186, 237, 238, 217]. The drawback of this approach, however, is that the
neutrino growth predicted by the fluid approximation can result in $\sim 100\%$ discrepancies
compared to Boltzmann codes. This feeds back as an error in the total matter growth
function (albeit with a much smaller amplitude that scales like f_ν) meaning that the initial
conditions will not be seeded in a manner that is consistent with the actual neutrino growth
predicted by the Boltzmann solver. Moreover, such large discrepancies in the neutrino growth
rate invalidate the use of the fluid approximation in checking whether or not the neutrino
particles end up growing correctly in the simulation.

We propose an alternative strategy that replaces the use of the fluid model with an
iterative scheme that assumes the neutrino growth relative to cold matter follows exactly
from the output of the Boltzmann code. The basic idea of our method is to replace the
value of Ω_m in the gravitational source term of the Poisson equation with a scale-dependent
quantity, $\Omega_m^{\mathrm{eff}}(k)$, defining the total amount of matter contributing to the growth of density
perturbations on scale k. We make use of the known asymptotic limits for growth in massive
neutrino cosmologies: on large scales neutrinos contribute to growth so that $\Omega_m^{\mathrm{eff}}(k \ll k_{\mathrm{fs}}) \to$
$\Omega_{cb\nu}$ while on small scales neutrinos free-stream meaning that $\Omega_m^{\mathrm{eff}}(k \gg k_{\mathrm{fs}}) \to \Omega_{cb}$. On
intermediate scales, we should observe a monotonic transition between the asymptotic limits.
This is then used to compute a scale-dependent growth factor, $D(z, \Omega_m^{\mathrm{eff}})$, that is calculated in
the same manner as is commonly done for massless neutrino cosmologies except that the Ω_m
term in the gravitational source term is replaced with $\Omega_m^{\mathrm{eff}}(k)$. The determination of $\Omega_m^{\mathrm{eff}}(k)$
is performed iteratively until we arrive at a self-consistent solution for all k. We circumvent
the use of approximate models for neutrino growth by interpolating from the direct output
of Boltzmann codes for the neutrino growth relative to cold matter. Of course, this method
is still approximate as we collapse the impact of the time-varying neutrino free-streaming
length into a single time-integrated quantity $\Omega_m^{\mathrm{eff}}(k)$ used to scale the growth between z_{ini}
and z_{fin}. However, as shown below, the errors introduced in this approximation are small
and notably improved compared to the commonly adopted approach.

The first step in each iteration of our scheme is to set up the initial densities and velocities
of the cold dark matter and baryons at the starting time, z_{ini}, of the simulation. The densities
are set by specifying $\delta_\alpha = T_\alpha^{\mathrm{bs}}$ following equation (3.5) with the growth factor evaluated
using the scale-dependent $\Omega_m^{\mathrm{eff}}(k)$ of that iteration. Note that only the Ω_m appearing in
the gravitational source term in the growth factor is replaced with Ω_m^{eff} and not any of the
background terms in the Hubble factor. The velocities are set by specifying $\theta_\alpha = T_{\theta,\alpha}^{\mathrm{bs}}$
following equation (3.7) with T_α^{bs} evaluated at $z_{\mathrm{ini}} \pm 0.1$ using the scale-dependent $D(z, \Omega_m^{\mathrm{eff}})$.
From here, we use a fourth-order Runge-Kutta method to evolve the coupled set of continuity,
Euler, and Poisson equations from $a_{\mathrm{ini}} = 1/(z_{\mathrm{ini}} + 1)$ to $a_{\mathrm{fin}} = 1/(z_{\mathrm{fin}} + 1)$ with steps taken
in $\ln a$:

$$\begin{aligned}
\frac{d\delta_{c,b}}{d\ln a} &= -\frac{\theta_{c,b}}{aH}, \\
\frac{d\theta_{c,b}}{d\ln a} &= -\left(\theta_{c,b} + \frac{3}{2}aH\Omega_m\delta_m\right),
\end{aligned} \tag{3.8}$$

where $\Omega_m = \Omega_{cb\nu}$ and $\delta_m = f_c\delta_c + f_b\delta_b + f_\nu\delta_\nu$ (with $f_\alpha \equiv \Omega_\alpha/\Omega_m$) is the total non-relativistic matter contribution at time a. Note that Ω_m in this expression is fixed as $\Omega_{cb\nu}$ in order to match how the gravitational forces are solved in the simulation. In other words, Ω_m^{eff} is used *only* in the calculation of the growth factor used to set the initial conditions and does not appear anywhere else in the iterative procedure.

Obviously, evaluating the expression above in a self-consistent manner requires knowledge of $\delta_\nu(a)$. The approach used in the REPS code of Ref. [228] is to couple the neutrinos in the integration of equation (3.8) with an additional neutrino pressure term that is modeled with the fluid approximation. Instead of attempting to directly model the growth of neutrino perturbations we rather interpolate from the output of the Boltzmann code. More specifically, we store the ratio $R_\nu = T_\nu/T_{cb}$ from CAMB at a series of scale factors between a_{ini} and a_{fin} (the ratio is deliberately used – over interpolating only T_ν – so as to cancel out the scale-dependencies associated with the relativistic contributions ignored in the Newtonain forward model). This is then used to set $\delta_\nu(a) = \delta_{cb}(a)R_\nu(a)$ where $\delta_{cb}(a)$ is evaluated at the instantaneous a of the time integrator and $R_\nu(a)$ is linearly interpolated from the sample points. In this way, we avoid the use of any approximations in the modeling of the neutrino pressure term. More importantly, this method is self-consistent since the ultimate goal of the simulation is to model the neutrinos in such a way that their growth relative to the total matter follows the linear theory predictions of the Boltzmann code at early times.

The first iteration in our scheme involves setting $\Omega_m^{\mathrm{eff}} = \Omega_{cb\nu}$ for all k in the backscaled growth factor. At the end of the first iteration, we find that the final solution for $\delta_m(z_{\mathrm{fin}}, k)$ matches with that expected from the unmodified CAMB density transfer function on large scales. On smaller scales, however, we find that $\delta_m(z_{\mathrm{fin}}, k)$ is suppressed since the initial density perturbations were backscaled under the incorrect assumption that neutrinos contribute to growth on those scales. At the end of the first iteration, we mark as sufficiently converged any k for which $\delta_m(z_{\mathrm{fin}}, k)$ is within $\epsilon = 10^{-6}$ of $T_m(z_{\mathrm{fin}}, k)$. In the second iteration, we set $\Omega_m^{\mathrm{eff}} = \Omega_{cb}$ for all remaining unconverged scales. The result at z_{fin} is the opposite of the first iteration: on small scales we find that δ_m matches $T_m(z_{\mathrm{fin}}, k)$ while on larger scales δ_m is enhanced with respect to CAMB since those scales were backscaled using an insufficiently small value for the gravitational source term. We again mark as converged any scales k that are within ϵ of the target solution. The result after two iterations is that the largest and smallest scales are converged while the results on intermediate scales can be linearly interpolated to make a guess for $\Omega_m^{\mathrm{eff}}(k)$ on the next iteration. This procedure is continued with each individual k updating its bracketing bounds of $\Omega_m^{\mathrm{eff}}(k)$ around either side of the root of the equation $y = \delta_m(z_{\mathrm{fin}}, k)/T_m(z_{\mathrm{fin}}, k) - 1$ until convergence is reached. In practice, we find that four iterations are usually sufficient to satisfy our convergence criteria for $\Omega_m^{\mathrm{eff}}(k)$.

We show in Figure 3.1 the converged result for $\Omega_m^{\mathrm{eff}}(k)$ for three different neutrino masses ($m_\nu = 0.05$, 0.15, and 0.45 eV) and three different initial redshifts ($z_{\mathrm{ini}} = 40$, 20, and 10). In each case we set $z_{\mathrm{fin}} = 0$ and use the cosmological parameters $(\Omega_c, \Omega_b, \Omega_\nu, \Omega_\Lambda, h, N_{\mathrm{eff}}) = (0.2684, 0.0491, m_\nu/(93.14h^2), 1 - \Omega_{cb\nu}, 0.6711, 2.046)$. For each case, we find the expected result that $\Omega_m^{\mathrm{eff}} \to \Omega_{cb\nu}$ (Ω_{cb}) on large (small) scales with a smooth transition between these limits on intermediate scales. The scales at which Ω_m^{eff} makes the transition between the two
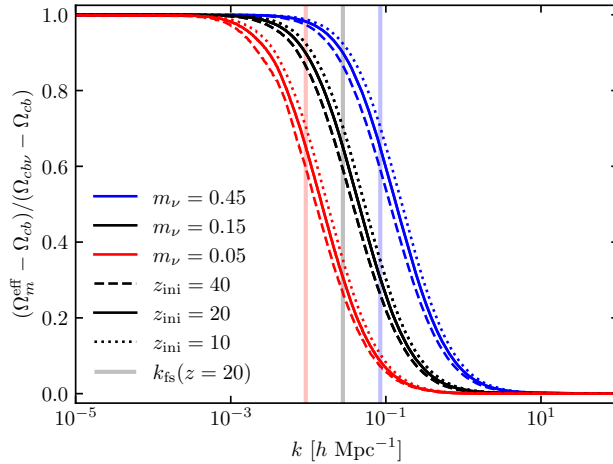
Figure 3.1: Converged $\Omega_m^{\mathrm{eff}}(k)$ for various neutrino masses and initial redshifts (normalized so that $\Omega_{cb\nu}$ is unity and Ω_{cb} is 0). For reference, the vertical shaded lines show the neutrino free-streaming scale evaluated at $z = 20$ for the corresponding neutrino mass.

asymptotic limits will roughly span the neutrino free-streaming scales covered between z_{ini} and z_{fin}. For reference, the vertical lines in Figure 3.1 show k_{fs} evaluated for each neutrino mass at $z = 20$. Note that the departure from the large-scale limit of $\Omega_m^{\mathrm{eff}} = \Omega_m$ occurs on larger scales as z_{ini} is increased with fixed neutrino mass. The reason is that $k_{\mathrm{fs}} \propto \sqrt{a}$ meaning that the inclusion of earlier times will reflect neutrino suppression on progressively larger scales. Hence, $\Omega_m^{\mathrm{eff}}(k)$ depends not only on m_ν, but also on the specific choices of z_{ini} and z_{fin}.

In Figure 3.2, we show the resulting backscaled total matter density transfer function, evaluated using the converged $\Omega_m^{\mathrm{eff}}(k)$ in equation (3.4), at various redshifts for the $m_\nu = 0.15$ eV case with $z_{\mathrm{ini}} = 40$. For reference, we show the backscaled transfer function divided by the unmodified CAMB transfer function evaluated at each redshift. By construction, our transfer function matches exactly with CAMB at $z_{\mathrm{fin}} = 0$ and deviates at higher redshift. In order to get a rough idea of the level of disagreement we expect with CAMB at high redshift, we also plot as dotted lines the ratio of backscaled transfer functions to CAMB for a massless neutrino cosmology. In this case, we use the same cosmological parameters as the massive neutrino cosmology but absorb Ω_ν into Ω_c and set $N_{\mathrm{eff}} = 3.046$. The backscaled transfer function in this case uses the scale-independent growth factor evaluated with $\Omega_m = \Omega_{cb}$. We do not expect the solid and dotted lines to exactly match, given that they do not correspond to the same cosmology and have differing levels of relativistic neutrino contributions, but we do observe qualitatively similar behavior in terms of how the backscaled transfer functions deviate from CAMB. In particular, we find that most of the discrepancy appears as an enhancement on large scales at early times. However, we do see slightly different behavior in the range of scales $10^{-2} \lesssim k/(\mathrm{Mpc}^{-1}h) \lesssim 10^{-1}$ with a small
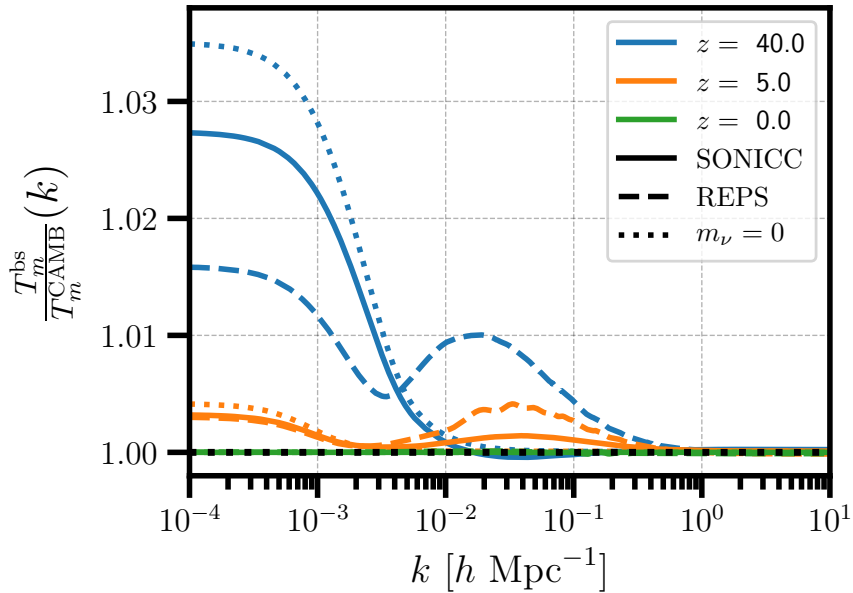
Figure 3.2: Total matter transfer functions from our iterative backscaling procedure (SONICC, solid lines) normalized to the unmodified CAMB transfer functions at various redshifts for a massive neutrino cosmology with $m_\nu = 0.15$ eV, $z_{\rm ini} = 40$, and $z_{\rm fin} = 0$. For comparison, the dashed lines show corresponding results from REPS while the dotted lines show results from a massless neutrino cosmology. For the latter, we use the same cosmological parameters as the massive neutrino case except that we absorb Ω_ν into Ω_c and also increase $N_{\rm eff}$ from 2.046 to 3.046.

enhancement (suppression) relative to the massless neutrino result at $z = 5$ (40). This difference reflects the approximate nature of integrating the impact of the time-varying neutrino free-streaming scale into the quantity $\Omega_m^{\rm eff}(k)$. Nevertheless, this error appears at only the $\sim 0.1\%$ level and is confined to only those scales in the known transition region around $k_{\rm fs}$.

To further illustrate the benefits of our approach, we add dashed lines in Figure 3.2 that correspond to the backscaled transfer functions evaluated using REPS for the same massive neutrino cosmology[3]. It is clear that the backscaled REPS transfer functions are behaving in a qualitatively different manner than both our iterative backscaling as well as the massless neutrino cosmology. In particular, the REPS transfer function exhibits $\sim 1\%$ enhancement on scales $10^{-2} \lesssim k/({\rm Mpc}^{-1}h) \lesssim 10^{-1}$ at $z = 40$ and seems to plateau to a different value on large scales than would be expected from comparison with the massless neutrino result. These features arise from the fluid approximation built into REPS which predicts neutrino growth that deviates from Boltzmann codes at the $\sim 100\%$ level (see Figure 3 of [228]).

---

[3]Following the choices made here, we use their "Scenario 4" method that assumes neutrino particles are simulated with a constant mass throughout their evolution.

This feeds back into the cold matter component, leading to the minor distortions seen in the effective growth factor for the total matter on intermediate scales. An additional drawback from this approach is that the inaccuracies of the fluid model in predicting neutrino growth make it difficult to check if the neutrinos end up growing correctly in the simulation. In this work, we use our backscaled transfer functions to perform a more careful analysis on how well the simulated neutrino growth matches expectations from the Boltzmann code.

We finish this section by noting that many massive neutrino simulations initialize the cold matter at an earlier redshift, z_{ini}^{cb}, than is used for the neutrinos, z_{ini}^{ν}. In fact, this is the default strategy used by HACC and employed in the simulations presented in the following sections. In this case, the backscaling method requires an additional modification in order to reflect the fact that the cold matter will grow in a scale-independent manner during the time $z_{\text{ini}}^{\nu} \leq z \leq z_{\text{ini}}^{cb}$. The first step is to further backscale the cold dark matter and baryon transfer functions obtained from the iterative method at z_{ini}^{ν} to the earlier starting time:

$$T_{\alpha}^{\text{bs}}(z_{\text{ini}}^{cb}, k) = F_{\alpha}(z_{\text{ini}}^{cb}, k) \frac{D(z_{\text{ini}}^{cb}, \Omega_m)}{D(z_{\text{ini}}^{\nu}, \Omega_m)} T_{cb}^{\text{bs}}(z_{\text{ini}}^{\nu}, k), \qquad (3.9)$$

where $T_{cb}^{\text{bs}}(z_{\text{ini}}^{\nu}, k)$ is obtained from the iterative backscaling procedure while $F_{\alpha}(z_{\text{ini}}^{cb}, k) \equiv T_{\alpha}(z_{\text{ini}}^{cb}, k)/T_{cb}(z_{\text{ini}}^{cb}, k)$ is taken directly from CAMB. Similarly, for the velocity transfer function we compute:

$$T_{\theta,\alpha}^{\text{bs}}(z_{\text{ini}}^{cb}, k) = F_{\theta,\alpha}(z_{\text{ini}}^{cb}, k) \frac{\dot{D}(z_{\text{ini}}^{cb}, \Omega_m)}{\dot{D}(z_{\text{ini}}^{\nu}, \Omega_m)} T_{\theta,cb}^{\text{bs}}(z_{\text{ini}}^{\nu}, k), \qquad (3.10)$$

where $T_{\theta,cb}^{\text{bs}}(z_{\text{ini}}^{\nu}, k)$ is obtained from the iterative backscaling procedure, the factor $F_{\theta,\alpha}(z_{\text{ini}}^{cb}, k) \equiv \dot{T}_{\alpha}^{\text{bs}}(z_{\text{ini}}^{cb}, k)/\dot{T}_{cb}^{\text{bs}}(z_{\text{ini}}^{cb}, k)$ with the dots denoting that we apply the finite difference method to the backscaled density transfer functions obtained with equation (3.9), and we define $\dot{D} \equiv a^2 H(\text{d}D/\text{d}a)$. Note that these equations preserve the familiar scale-independent relations $T_{cb} \propto D$ and $T_{\theta,cb} \propto \dot{D}$ seen for the cold matter field in massless neutrino cosmologies.

It is important to note that the Ω_m used in equations (3.9) and (3.10) is set to $\Omega_{cb\nu}$ for all scales even though the neutrinos are not simulated above z_{ini}^{ν}. In order for the simulation to achieve consistent growth, this convention requires that Ω_{ν} be absorbed into the cold matter particle masses for $z > z_{\text{ini}}^{\nu}$. We have chosen this setup since it guarantees that the cold matter densities and velocities after evolution to z_{ini}^{ν} end up matching with the backscaled transfer functions previously computed at z_{ini}^{ν}, but only on the large scales for which $\Omega_m^{\text{eff}} \simeq \Omega_{cb\nu}$. On smaller scales, we incur an inconsistency with the cold matter densities (velocities) being slightly suppressed (enhanced) at the sub-percent level compared to the $T_{\alpha}^{\text{bs}}(z_{\text{ini}}^{\nu}, k)$ $[T_{\theta,\alpha}^{\text{bs}}(z_{\text{ini}}^{\nu}, k)]$ previously computed from the iterative method. This is not an issue since the total matter still converges to the same answer at z_{fin}, but does mean that we end up with a sub-percent inconsistency in the initial neutrino component as it is set using the $T_{\nu}^{\text{bs}}(z_{\text{ini}}^{\nu}, k)$ that assumed slightly different cold matter density and velocity amplitudes on small scales. An alternative approach would be to set $\Omega_m = \Omega_{cb}$ in equations (3.9) and (3.10) without changing the simulation particle masses between z_{ini}^{cb} and z_{ini}^{ν}, but this would just mean that we incur a minor inconsistency on large scales instead of small scales. In

general, starting the cold matter at an earlier time than the neutrinos requires that we incur an inconsistency somewhere due to the discontinuous nature of transitioning from scale-independent to scale-dependent growth at z_{ini}^{ν}. Pushing this inconsistency to small scales is preferable since these will be the first to undergo non-linear growth that tend to wash out small errors in the initial conditions. Either way, we stress that the inconsistency manifests at the sub-percent level in the neutrino component meaning that its impact on the total matter is exceedingly small.

For convenience, we have created a public code named SONICC[4] (Scale-dependent Omega for Neutrino Initial Condition Codes) that uses the iterative backscaling procedure presented above to generate initial density and velocity transfer functions for cold dark matter, baryons, and neutrinos for an arbitrary choice of cosmological parameters as well as starting and stopping redshifts (including the ability to start the cold matter earlier than neutrinos). The code is written in Python3 and uses the public `camb` module to compute Boltzmann transfer functions at the initial and final times as well as interpolate the relative neutrino growth between these endpoints. It was shown in Ref. [202] that the neutrino transfer functions produced by CAMB are sensitive to the choice of accuracy settings `accuracy_boost` and `l_accuracy_boost`. Our default choice is to take `accuracy_boost` $= 4$ and `l_accuracy_boost` $= 5$, which means that the neutrino transfer functions are accurate at the few-percent level for $k \gtrsim 0.1\ h\ \mathrm{Mpc}^{-1}$. With these accuracy settings, the typical runtime of the backscaling code is about 8 minutes, with the bulk of this coming from the computation of CAMB transfer functions. This can be reduced to under a minute by decreasing the default accuracy settings, but doing so comes at the expense of decreased accuracy in the neutrino transfer functions as well as the resulting $\Omega_m^{\mathrm{eff}}(k)$.

3.3   Neutrino Initial Conditions and Evolution

Theoretical modeling of the growth of LSS in the presence of massive neutrinos will require an equal or better level of precision as that of observational measurements. Cosmological simulations of structure formation are vital in this regard. In this work, we extend the code HACC [229] to evolve neutrinos as an additional N-body particle species alongside the cold matter. HACC is a cosmological simulation code designed to run performantly with extreme computational loads on all modern supercomputing systems. This performance is inherited here since the inclusion of neutrino particles mainly involves modifications to the initial conditions while preserving the core code structure and algorithmic choices. Recently, a hydrodynamic extension to HACC, known as CRK-HACC [239], was developed and we plan to merge the work presented here with the hydrodynamic extension so that fully self-consistent three-species (cold dark matter, baryons, neutrinos) simulations can be performed efficiently at scale.

There are two main challenges associated with the cosmological simulation of neutrino particles: 1) the ambiguities associated with properly backscaling massive neutrino cosmolo-

---

[4]https://git.cels.anl.gov/hacc/sonicc

gies in a way that is consistent with the Newtonian forward model; 2) dealing with the high
thermal motion that tends to randomize neutrino particles across the simulation domain. To
overcome the first issue, we use the iterative backscaling procedure presented above to cre-
ate more accurate initial transfer functions that are valid for all three non-relativistic matter
species. For the second issue, we make use of the "tiling" procedure [224], which utilizes
symmetries in the neutrino thermal momentum sampling in order to significantly reduce
the level of shot noise associated with randomization of the neutrino particles. We briefly
describe this method below and expand on the minor modifications to its implementation
that were used in this work.

In the standard picture, the cosmic neutrino background (CNB) constitutes a thermal
relic species from the early universe. More specifically, massive neutrinos decoupled from
the early plasma when the universe was roughly one second old and remained free of all in-
teractions save gravity until the present day. During this time, the temperature of the CNB
cooled with the expansion of the universe and has a current temperature $T_{\nu,0} \simeq 1.95$ K. At
early times, the neutrinos behave as a relativistic radiation component while later tran-
sitioning into a non-relativistic matter component when the temperature-to-mass ratio,
$(k_B T_{\nu,0})/(a m_\nu)$, drops below $\mathcal{O}(1)$. The massive neutrino comoving momentum magnitudes,
q, are set by the relativistic Fermi-Dirac distribution that describes the neutrino phase space
after decoupling:

$$f_0(q) = \frac{g_s}{2\pi^2 \hbar^3} \frac{q^2}{e^{\frac{q}{k_B T_\nu}} + 1}, \tag{3.11}$$

where $q \equiv ap$, with p the neutrino momentum magnitude and g_s, the spin factor.

It is common in cosmological simulations to initialize neutrino particles in an analogous
manner to cold matter. First, the neutrino particles are placed on a uniform lattice containing
N_ν^3 points with gravitationally-induced displacements and peculiar velocities computed from
the initial transfer functions. However, unlike the cold matter, a thermal velocity is also
added to each neutrino particle with a random direction vector and random magnitude drawn
from the Fermi-Dirac distribution. Occasionally, this stage is modified so that each lattice
site initializes a pair of neutrino particles with the pair having random thermal velocities
equal in magnitude but opposite in direction [204]. In any event, the main drawback of
this method is that the thermal velocities are so large that using a random draw results
in a density field that is essentially a Poisson sampling of the spatial neutrino distribution,
resulting in a shot noise term in the power spectrum that is inversely proportional to the
total number of neutrino particles, $P_{\text{shot}} \propto N_\nu^{-3}$. Since free-streaming suppresses neutrino
perturbations on small scales, a prohibitively large number of particles may be required to
resolve the intrinsic neutrino signal below the shot noise level.

In the "tiling" method [224], the shot noise is reduced by imposing symmetries on the
initial neutrino velocity distribution. To begin, this method preserves the use of a neutrino
grid containing N_ν^3 points, but associates a total of $N_{\text{SHELL}} \times N_{\text{DIR}}$ particles to each grid
point. This collection of neutrino particles is viewed as a set of N_{SHELL} independent spherical
shells that each contain N_{DIR} particles. The thermal velocity magnitude is varied across the

different shells, but held fixed for all particles within the same shell. Hence, only a total of N_{SHELL} unique velocity magnitudes are used for the thermal velocities and these are chosen in such a way as to optimize the sampling of the full Fermi-Dirac distribution. Within each shell, the velocity directions are chosen by taking a symmetric discretization of N_{DIR} points on the unit sphere. In the original implementation, these direction vectors are set using HEALPix [240], though other variations including minimizing the Coulomb potential of point charges on a sphere [213] and utilizing a Fibonacci grid [215] yield similar results. The crucial aspect of the tiling method is that this discretization of the velocity magnitudes and directions is replicated at each of the N_ν^3 grid points so that a constant flux of neutrino particles pass through adjacent volumes in the simulation domain, thereby suppressing the shot noise.

In the HACC simulations presented here, our general setup is to initialize a grid of N_{cb}^3 cold matter particles with an N_ν^3 neutrino grid that is maximally offset from the cold matter grid in order to reduce the potential impact of artificial particle coupling [241]. We split the Fermi-Dirac distribution into N_{SHELL} bins of equal probability mass and associate the mean value in each bin to the neutrino shells whose limits $\{q_i^{\min}, q_i^{\max}\}$ are determined by the inverse Fermi-Dirac cumulative distribution function:

$$q_i = \sqrt{\frac{\int_{q_i^{\min}}^{q_i^{\max}} dq\, q^2 f_0(q)}{\int_{q_i^{\min}}^{q_i^{\max}} dq\, f_0(q)}} \tag{3.12}$$

In the simulations presented here, we use HEALPix to assign the N_{DIR} initial momentum directions. We have also experimented with the use of a Fibonacci grid, but did not observe any significant difference compared to HEALPix for a fixed number of directions. The HEALPix directions are parameterized by the internal variable N_{SIDE} which gives, e.g., $N_{\mathrm{DIR}} = (12, 48, 192)$ for $N_{\mathrm{SIDE}} = (1, 2, 4)$. Motivated by the results of Ref. [213], we modify the original tiling scheme so that the direction vectors are rotated between shells. In particular, for shell i, we rotate the N_{DIR} direction vectors by an amount $(i/N_{\mathrm{SHELL}})(\pi/2)$ about the symmetry axis of the HEALPix discretization (which we align with the z axis). As shown later, this method helps to mitigate numerical effects associated with the tiling scheme symmetries, particularly for smaller N_{SIDE}. In the recent work of Ref. [215], the authors advocate for radially displacing each shell from the grid site by an amount proportional to its momentum magnitude in order to reduce a spurious coupling with the cold matter. We do not use this method here (meaning that each shell is sourced at the same grid location), but rather apply a spectral filter in the force solver, discussed below, in order to minimize the back-reaction of the neutrino grid on the cold matter.

The initial displacements and (non-thermal) velocities of the cold matter and neutrino particles are computed using the Zel'dovich approximation (ZA) evaluated on an N_{cb}^3 grid with the backscaled transfer functions of each respective species. The same random phases are used for each species with a rotation applied in Fourier space for the neutrino particles to account for their real-space offset from the cold matter grid where the displacement field is

sampled. We only consider the (unphysical) case of a single massive neutrino here, but plan to accommodate multiple mass eigenstates, and the complexities therein, in future work. By default, we initialize the cold matter at $z_{\rm ini} = 200$ while starting the neutrinos at a lower redshift $z^\nu_{\rm ini}$ (we consider $z^\nu_{\rm ini}$ within a range of 10-40). There are competing effects to consider when choosing the starting neutrino redshift, and the values used here are chosen as a compromise between the two limiting regimes. On the one hand, starting neutrinos too early is complicated by the fact that a significant fraction will still be relativistic – meaning that both their initialization as a non-relativistic species as well as their subsequent evolution will be incorrectly modeled within the Newtonian dynamics of the simulation [242]. On the other hand, starting neutrinos too late means that they become less well-described by linear theory, particularly for the slower-moving population. One way of pushing the starting redshift lower would be with the use of higher-order Lagrangian perturbation theory for initializing both species [243] (rather than just the ZA, or 1LPT). The impact of starting at lower redshift using higher-order perturbation theory has been studied by several groups [e.g., 244, 245]; however, a fully self-consistent multi-species approach is not currently available.

The gravitational force calculation in HACC is decomposed into two parts: (1) the long-range force evaluated using a spectral particle-mesh (PM) method that solves the (filtered) Poisson equation in Fourier space; (2) the short-range force evaluated using direct pairwise interactions with an algorithm optimized for the given hardware (i.e., we employ a particle-particle method on accelerated systems and a tree method on CPU platforms). The two components are delineated by the force-matching scale, r_s, that is roughly equal to three times the length of an individual cell in the PM mesh. Our default choice is to use a PM mesh that is of the same size as the initial cold matter particle grid. The main time step is governed by the PM force evaluation while the short-range force is evaluated on a shorter "subcycled" time-scale. For the simulations presented here, we take 625 PM steps between $z_{\rm ini}$ and $z_{\rm fin}$ with 4 short-range force evaluations per PM step, following the typical integration procedure used in HACC gravity-only simulations [229].

The force decomposition and time-stepping remain unchanged with neutrinos except for two minor modifications. Firstly, we have added the flexibility to exclude neutrino particles from the short-range force calculation. This is simply achieved by sorting neutrinos to the end of the particle arrays and only passing the cold matter portion of those arrays to the short-range kernel. This reduces the computational cost of the short-range force calculation, but implicitly assumes that the gravitational impact of neutrinos below the force-matching scale can be safely ignored. We use this method for the simulations presented here, which we justify by the fact that the force-matching scale $r_s \sim 1\ h^{-1}$Mpc while the free-streaming scale $1/k_{\rm fs} \gtrsim 10\ h^{-1}$Mpc for the neutrino mass and box sizes considered here. The second modification is that we have implemented an optional filter that smooths the neutrino density field in Fourier space during the long-range PM force calculation. This is achieved using a simple sharp-k smoothing filter applied in Fourier space via the transformation:

$$\rho^{\rm sm}_\nu(\mathbf{k}) = \Theta(k_{\rm cut} - |\mathbf{k}|)\rho_\nu(\mathbf{k}), \tag{3.13}$$

where $\rho_\nu(\mathbf{k})$ is the Fourier transform of the neutrino density field, computed in real space

using a cloud-in-cell (CIC) interpolation to the PM mesh, and Θ is the Heaviside step function with $k_{\text{cut}} = \pi N_\nu / L$ chosen to be the Nyquist frequency, k_{Nyq}^ν, of the neutrino grid. As shown in Section 3.4, this is done to avoid potential imprinting of the neutrino grid onto the cold matter density field, which we have found to occur if the neutrino mass resolution is sufficiently coarse. Again, this smoothing assumes that neutrinos are not significantly clustering below the cutoff scale, which we justify since $k_{\text{Nyq}}^\nu \gtrsim 10 k_{\text{fs}}$ for the main runs presented here.

The final consideration that we make is in regards to the particle "overloading" performed in HACC. In general, the global simulation volume is subdivided across MPI ranks with each rank assigned a subvolume with side lengths on the order of a few $10 - 100\ h^{-1}$Mpc along each axis. This "alive" zone of each rank is then extended $2 - 10\ h^{-1}$Mpc at the edge of each boundary and filled with particle replicants from the alive zones of neighboring ranks. This is done to ensure proper boundary conditions are utilized in the short-range force for particles at the edge of the alive zone. The overload zone is refreshed at regular intervals (typically every 1-20 PM steps) with the cadence chosen as a balance between minimizing the MPI communication overhead and the propagation of gravitational force errors from the edge of the overload zone inwards to the alive zone (see Ref. [229] for more details).

One of the challenges with neutrino simulations is that the higher momentum neutrinos are still moving relatively fast near the initialization redshift. The problem is that if a neutrino particle is capable of moving further than the width of the overload zone during one refresh cycle then it is possible that all of its replicants end up outside the alive zone at the next refresh. The outcome is that the particle will be removed from the simulation and we may end up carving out neutrinos near the boundaries of each MPI rank. The trivial solution is to increase the refresh rate and/or increase the overload zone. For the simulations presented here, we use a refresh every PM step and increase the overload to $20\ h^{-1}$Mpc at early times. This choice adds a significant computational overhead and in the future we plan to avoid large overloads by implementing a method to manually pass neutrino particles across MPI ranks when they would otherwise be removed from the simulation.

3.4 Results

We present here results from a suite of HACC runs that vary the internal parameters of the neutrino tiling scheme as well as the simulation resolution. The full set of runs are enumerated in Table 3.1. In addition to varying the two main parameters of the tiling scheme – the number of momentum shells N_{SHELL} and the number of directions per shell N_{DIR} (as set by the HEALPix parameter N_{SIDE}) – we also vary the size of the neutrino grid N_ν. This allows us to check for numerical convergence in the power spectra as one parameter is varied while the others are held fixed. The main set of runs use a simulation box of width $L = 250\ h^{-1}$Mpc containing $N_{cb}^3 = 512^3$ cold matter particles though we also consider two other runs with a larger box of $L = 1000\ h^{-1}$Mpc. We also explore the impact of the neutrino starting redshift by varying $z_{\text{ini}}^\nu = (10, 20, 40)$ with z_{ini}^{cb} held fixed at 200.

Name	N_{cb}	N_ν	N_{SHELL}	N_{DIR}	z^ν_{ini}	Filter	L	Rotate ϕ
`fid_nophi`	512	128	5	12	20	Yes	250	F
`fid_nofil`	512	128	5	12	20	No	250	T
`fid`	512	128	5	12	20	Yes	250	T
`med_N`ν	512	256	5	12	20	Yes	250	T
`hi_N`ν	512	512	5	12	20	No	250	T
`med_NDIR`	512	128	5	48	20	Yes	250	T
`hi_NDIR`	512	128	5	192	20	Yes	250	T
`med_NSH`	512	128	10	192	20	Yes	250	T
`hi_NSH`	512	128	20	192	20	Yes	250	T
`coarse_hi_NDIR`	512	64	5	192	20	Yes	250	T
`lo_z`	512	128	5	192	10	Yes	250	T
`hi_z`	512	128	5	192	40	Yes	250	T
`lb_loN_nofil`	256	64	5	12	40	No	1000	T
`lb_loN`	256	64	5	12	40	Yes	1000	T

Table 3.1: Simulation suite considered in this work. All simulations initialize the cold matter at $z^{cb}_{\text{ini}} = 200$ in a box of side length L while the neutrinos are initialized later, at z^ν_{ini}. In each case, we use the cosmological parameters, $(\Omega_c, \Omega_b, \Omega_\nu, \Omega_\Lambda, \sigma_8, h, N_{\text{eff}}) = (0.2684, 0.0491, m_\nu/(93.14h^2), 1\text{-}\Omega_{cb\nu}, 0.8, 0.6711, 2.046)$, for a single massive neutrino species with $m_\nu = 0.15$ eV. The cold matter contains N^3_{cb} particles while the total number of neutrino particles is given by the product $N^3_\nu \times N_{\text{SHELL}} \times N_{\text{DIR}}$. In each run, the long-range force is calculated using a PM mesh containing $N_g = N_{cb}$ cells per side; the short-range force is evaluated only on the cold matter and uses a Plummer softening length of $0.1L/N_g$. The "Filter" column denotes whether the sharp-k force filter in equation (3.13) is applied to the neutrino density field during the long-range force evaluation. The "Rotate ϕ" column indicates whether the momentum shells were individually rotated, as described in Section 3.3.

In Section 3.4 we present a numerical convergence study for the neutrino tiling parameters. We then show in Section 3.4 that one challenge with the tiling method is that the Fourier imprint of the neutrino grid can transfer onto the cold matter power spectrum if the simulation resolution is sufficiently coarse. We proceed in Section 3.4 with a comparison of the simulated total matter power spectrum to emulator predictions [194] and finish in Section 3.4 with a closer look at the linear theory growth of neutrino power. In all of the further analysis, we examine the power spectra of the cold matter and neutrino components which are computed using a CIC interpolation of each species onto a mesh containing 512^3 cells.

Convergence Study of Neutrino Initialization Parameters

We begin by varying N_{DIR} with all other parameters held fixed. In Figure 3.3 we show the cold matter and neutrino power spectra at redshifts $z = 9$, 1, and 0 for the `fid_nophi`, `fid`, `med_NDIR`, and `hi_NDIR` runs. All runs contain $N_{cb}^3 = 512^3$ cold matter particles with an $N_\nu^3 = 128^3$ neutrino particle grid consisting of $N_{\mathrm{SHELL}} = 5$ momentum shells. In addition, all of the runs except for `fid_nophi` use the method described in Section 3.3 of rotating the direction vectors of each shell. Comparing the dashed orange and dot-dashed green lines in Figure 3.3 show that this rotation has a significant impact on the neutrino power spectrum, both at early and late times. This behavior was observed previously [213] and likely reflects the fact that the rotation reduces the correlation in neutrino particle trajectories of adjacent momentum shells. Though not shown here, we find that the relative impact of the rotation decreases as we increase N_{DIR}. This can be attributed to the fact that the direction vectors become increasingly isotropic as the discretization is made increasingly fine.

Comparing the dot-dashed green, dotted red, and solid blue lines in Figure 3.3 shows that the neutrino power spectrum is quite sensitive to the number of direction vectors. At $z = 9$, the $N_{\mathrm{DIR}} = 12$ and 48 runs are converged to the $N_{\mathrm{DIR}} = 192$ run at the 10% level up to $k \simeq 0.1$ and 0.3 h Mpc^{-1}, respectively. This improves at $z = 0$ with the 10% agreement extending to $k \simeq 0.3$ and 0.7 h Mpc^{-1} for the two runs. Hence, the $N_{\mathrm{DIR}} = 192$ run appears to be reasonably converged (at the $\sim 10\%$ level) up to half the Nyquist frequency of the neutrino grid at $z = 0$, though more directions may be required to reach a similar level of convergence at higher redshift. Note, however, that as we increase N_{DIR}, the neutrino power spectrum on scales $k \gtrsim 0.1$ h Mpc^{-1} converges to a result at $z = 9$ that is markedly below the linear theory curve; we will discuss this discrepancy in Section 3.4. The large differences in neutrino power spectra for $k > k_{\mathrm{Nyq}}^\nu$ simply reflect different levels in the effective neutrino shot noise of each run. The shot noise level is governed by the total particle count, as shown by the horizontal line tracing $P_{\mathrm{shot}} = L^3/(N_\nu^3 N_{\mathrm{SHELL}} N_{\mathrm{DIR}})$ for the `hi_NDIR` run. This is also the level at which the neutrino power spectrum saturates to in standard simulations that draw random velocity magnitudes and directions. The fact that we are able to resolve the neutrino power spectrum below P_{shot} for $k \leq k_{\mathrm{Nyq}}^\nu$ thus confirms the utility of the tiling method, as also verified in Refs. [224, 215, 213]. Note that the neutrino power spectrum still saturates to P_{shot} for $k > k_{\mathrm{Nyq}}^\nu$ with the presence of noticeable spikes observed at the Fourier modes of the neutrino grid. We will explore this topic in more detail in Section 3.4.

Despite the large variance observed in neutrino power spectra with varying N_{DIR}, we see that the cold matter power spectra are relatively unchanged between each run. This is particularly true at high redshift and linear scales though we do observe a $\sim 0.5\%$ scatter near the cold matter Nyquist frequency at $z = 0$. As further shown below, we generally find that the neutrino power spectra exhibit systematic numerical convergence as one parameter is varied while all others are held fixed. The cold matter power spectra, on the other hand, mainly show small random noise on non-linear scales. We attribute this to the fact that minor changes in the neutrino distribution are not strong enough to significantly alter the cold matter on linear scales (since f_ν is small), but are able to seed noise, particularly in the
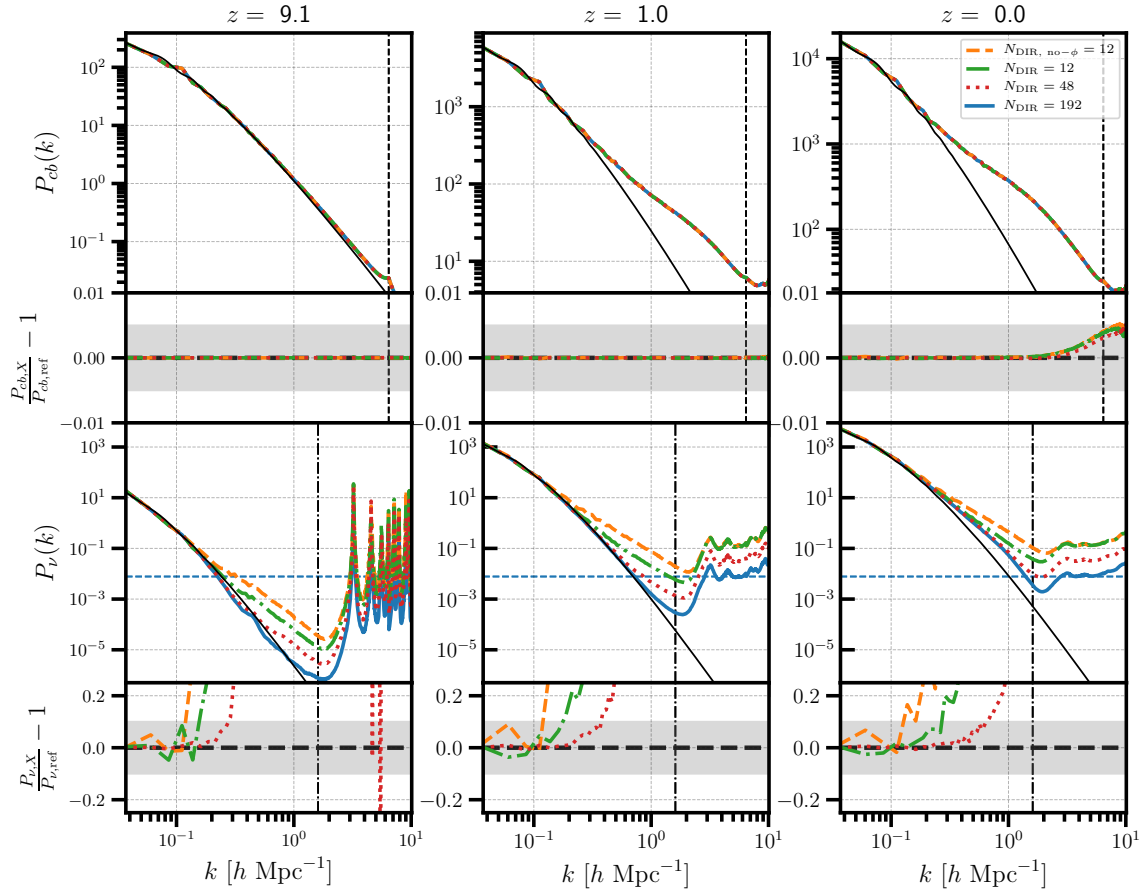
Figure 3.3: Cold matter (upper two rows) and neutrino (lower two rows) power spectra for different choices of number of neutrino particle velocity directions (`fid` [green dash-dotted], `med_NDIR` [red dotted], `hi_NDIR` [blue solid]), as well as without rotating HEALPix momentum shells (`fid_nophi` [orange dashed])). The ratio of the power spectrum at each choice of parameters relative to the power spectrum of the `hi_NDIR` run ($N_{\mathrm{DIR}} = 192$) is shown in the residual panels (second and fourth rows). Shaded bands in the residual panels are at 0.5% for cold species and 10% for neutrinos. Plots are shown at several redshifts in the different columns. Backscaled linear theory is also shown for reference as the thin black curves. In the third row of panels, the dashed blue line corresponds to the expected Poisson shot noise for the `hi_NDIR` run which contains the largest neutrino particle count. The vertical dashed black line in the upper panels is the Nyquist mode of the initial cold species grid, while the dash-dotted black line in the lower panels is the same for the initial neutrino grid.

shot-noise dominated regime, that becomes amplified with non-linear growth. Given that the noise is random in nature (i.e., it fluctuates between runs in a non-systematic manner) and confined to highly non-linear scales, its presence does not indicate that the tiling scheme

fails to numerically converge, rather that care must be taken to understand the amplitude of random error in the total matter distribution on small scales.

Next we consider the case where we vary N_{SHELL} while keeping the other parameters fixed. Similar to Figure 3.3, Figure 3.4 shows cold matter and neutrino power spectra at redshifts $z = 9, 1$, and 0 for the `fid`, `med_NSH`, and `hi_NSH` runs, which have $N_{\mathrm{SHELL}} = 5, 10, 20$ respectively (all at $N_{\mathrm{DIR}} = 192^5$). On the largest scales we find good agreement, with all three runs following linear theory closely. On intermediate scales, the neutrino power in the three runs starts to diverge at around $k \approx 0.2 \ h \ \mathrm{Mpc}^{-1}$ at the 10% level. As expected, the $N_{\mathrm{SHELL}} = 10$ neutrino power agrees more with that of the $N_{\mathrm{SHELL}} = 20$ run than the $N_{\mathrm{SHELL}} = 5$ run does, and this is true at all redshifts. For $k > k_{\mathrm{Nyq}}^{\nu}$, the difference between the neutrino power spectra is due to a change in shot noise since the $N_{\mathrm{SHELL}} = 20$ (10) run has four times (twice) as many neutrino particles as the $N_{\mathrm{SHELL}} = 5$ run. Meanwhile, the cold species power is essentially unaffected by the change in N_{SHELL} at the 0.1% level for linear scales. At later times, we again see fluctuations in the cold matter power on the non-linear scales near the Nyquist mode. The random nature of this noise is evident in the flipping and reordering of the green and orange curves when comparing the $z = 1$ and $z = 0$ residual panels. Overall, we find a less dramatic effect on the power spectra of both species when varying the number of shells compared to the number of directions; in agreement with the initial exploration of Ref. [224].

Having characterized how the cold species and neutrino power spectra change with N_{DIR} and N_{SHELL}, we now consider varying the resolution of the neutrino grid N_{ν} in Figure 3.5. Following our earlier convention, we show the cold matter and neutrino power spectra for several values of N_{ν} in different colors, where solid (dashed) lines correspond to runs with $N_{\mathrm{DIR}} = 12$ (192) and $N_{\mathrm{SHELL}} = 5$ held fixed in each case. Since all simulations have $N_{cb} = 512^3$, we can consider this analysis to represent making different choices of the ratio N_{ν}/N_{cb}. In terms of the neutrino power spectra, the general trend of increasing N_{ν}/N_{cb} is independent of the value of N_{DIR}. Namely, we find that the neutrino power spectrum is insensitive to the choice of N_{ν}/N_{cb} up until the Nyquist frequency of the neutrino particle grid. This is easily seen in the bottom panel, where each line traces a residual of zero up until the vertical dotted line indicating its Nyquist wavenumber. As mentioned before, comparing the two sets of N_{DIR} runs show that the neutrino power spectrum is significantly enhanced on intermediate scales for $N_{\mathrm{DIR}} = 12$. Furthermore, Figure 3.5 shows that up until a comoving wavenumber of $k \approx 3.0 \ h \ \mathrm{Mpc}^{-1}$ (at $z = 0$), the $N_{\nu} = 256$ and $N_{\nu} = 512$ neutrino power spectra agree at the 10% percent level. This comparison is important for considering the computational cost of the tiling method, as the former run has a factor of 8 fewer neutrino particles than the latter. We see the same general behavior in the cold matter power as was shown in the previous two tests. Namely, the cold matter power displays random noise on small non-linear scales. As before, the cold matter is relatively insensitive to changes in N_{ν}/N_{cb} on linear scales.

---

[5]We also found qualitatively similar results when varying N_{SHELL} when the number of directions was fixed to $N_{\mathrm{DIR}} = 12$.
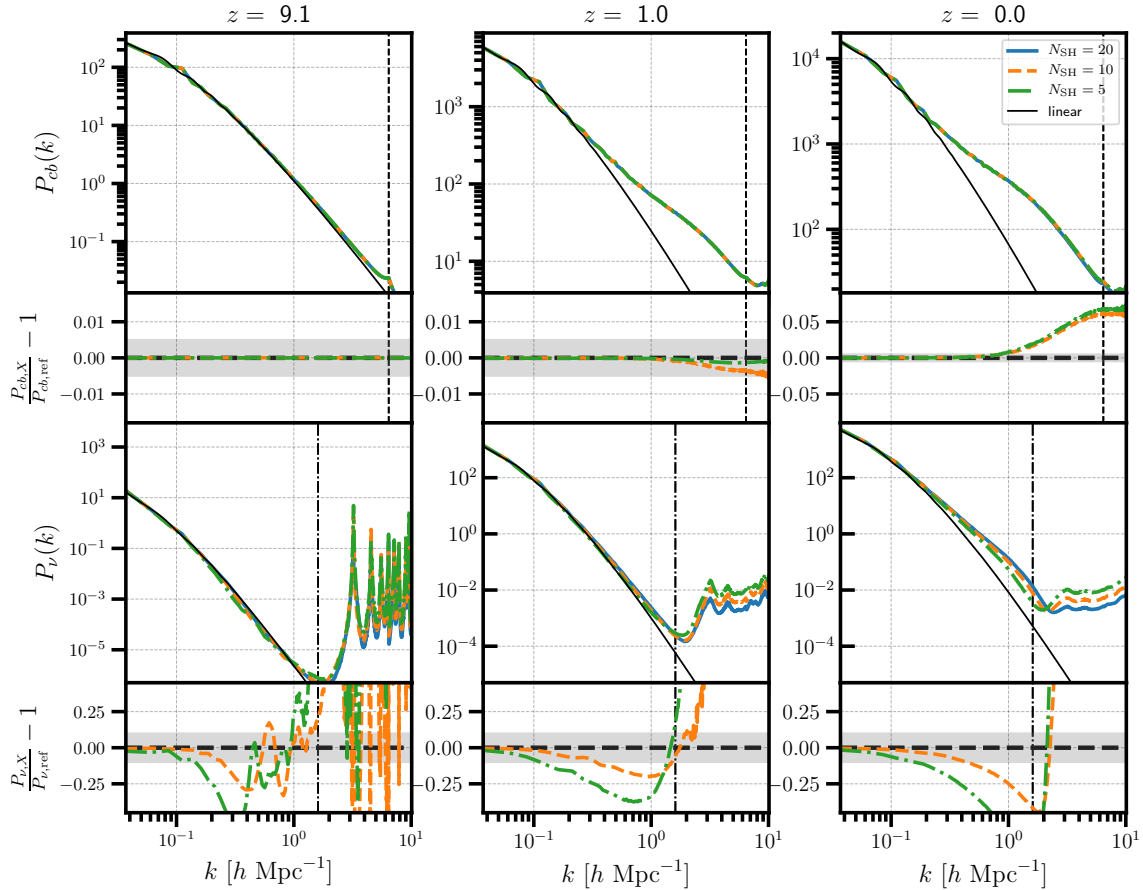
Figure 3.4: Cold matter (upper two rows) and neutrino (lower two rows) power spectra for different choices of number of neutrino momentum bins (`fid` [green dash-dotted], `med_NSH` [orange dashed], `hi_NSH` [blue solid]). All simulations have $N_{\mathrm{DIR}} = 192$. The ratio of the power spectra with respect to the `hi_NSH` run with $N_{\mathrm{SHELL}} = 20$ is shown in the residual panels (second and fourth rows). Shaded bands in the residual panels are at 0.5% for cold species and 10% for neutrinos. Plots are shown at several redshifts in the different columns. The vertical dashed black line in the upper panels is the Nyquist mode of the initial cold species grid, while the dash-dotted black line in the lower panels is the same for the initial neutrino grid. Backscaled linear theory is also shown for reference as the thin black curves.

The results presented above show that numerical convergence at the $\sim 10\%$ level in the neutrino power spectra requires $N_{\mathrm{DIR}} \gtrsim 100$ direction vectors and $N_{\mathrm{SHELL}} \gtrsim 10$ momentum shells. The cold matter power is not significantly impacted by choices in these parameters on linear scales, but does exhibit random noise at the $\sim 1\%$ level on non-linear scales at late times. We suspect that this error is seeded by shot noise from the neutrinos on small scales (which scales inversely with the total number of neutrino particles) and becomes amplified

by non-linear growth. The main challenge with the tiling method is that the requirement $N_{\rm DIR} \gtrsim 100$ and $N_{\rm SHELL} \gtrsim 10$ implies that $N_\nu/N_{cb} \lesssim 1/10$ for the number of neutrino particles to not significantly exceed the cold matter count. The coarseness of the neutrino particle grid limits the smallest scale at which neutrinos are effectively resolved and, as shown below, requires extra care to prevent discreteness effects from artificially back-reacting on the cold matter.

Suppression of Neutrino Grid Artifacts

The runs shown in Figure 3.5 display characteristic small-scale spikes in the neutrino power spectra with locations that depend on the value of N_ν. These spikes were also observed in Ref. [213] and correspond to resonances in the neutrino particle grid which are picked up as essentially delta functions in Fourier space. More specifically, these spikes occur at multiples of the Nyquist frequency of the neutrino particle grid (e.g., the first two spikes occur at $2k^\nu_{\rm Nyq}$ and $2\sqrt{2}k^\nu_{\rm Nyq}$). Of course, this is not a feature unique to the neutrino particles; if we were to measure the cold matter power spectrum on a fine enough mesh then we would likewise observe a series of spikes at the Fourier modes corresponding to the cold matter particle grid [246]. In other words, single-species N-body simulations also contain the Fourier imprint of the initial conditions grid that are subsequently captured in the force solver provided the force resolution is finer than the particle grid separation. These grid features are generally considered to insignificantly impact evolution in the single-species case due to the efficient gravitational transfer of power from large to small scales [e.g., 247]. The result is that gravity tends to broaden the spikes in Fourier space until the gravitational growth on small scales surpasses the shot noise level, erasing the spikes altogether. This gravitational broadening can be seen in the lower panels of Figure 3.5 for the neutrinos and is also examined more extensively in Ref. [213]. However, it is not clear what impact these features may have in multi-species simulations especially if the Nyquist frequencies of the particle grids are mismatched, as is the general case here.

To study this issue more closely, we ran two simulations with $N_\nu < N_{cb}$ and varying mass resolutions to test whether we could clearly identify the imprint of the neutrino grid on the cold matter. Indeed, we found that for sufficiently coarse resolutions, it is possible for the neutrino grid to transfer its Fourier signal onto the cold matter power spectrum. This is demonstrated in Figure 3.6 where we show the power spectrum at $z = 5$ for the `lb_loN_nofil` simulation that contains $N^3_{cb} = 256^3$ cold matter particles with an $N^3_\nu = 64^3$ neutrino particle grid in an $L = 1000~h^{-1}{\rm Mpc}$ box. First, we see the characteristic spikes in the neutrino power spectrum which extend prominently above the neutrino shot noise level. Furthermore, we also clearly observe a small spike in the cold matter power spectrum situated at the location of the first neutrino spike, as shown in the plot inset. The presence of this spike unambiguously demonstrates that the numerical imprint of the neutrino particle grid is being transferred onto the cold matter density field via the force solver. We checked for similar spikes in the cold matter power spectra of two other simulations, `fid_nofil` and `fid`, but did not observe any, indicating that this imprinting depends on the simulation
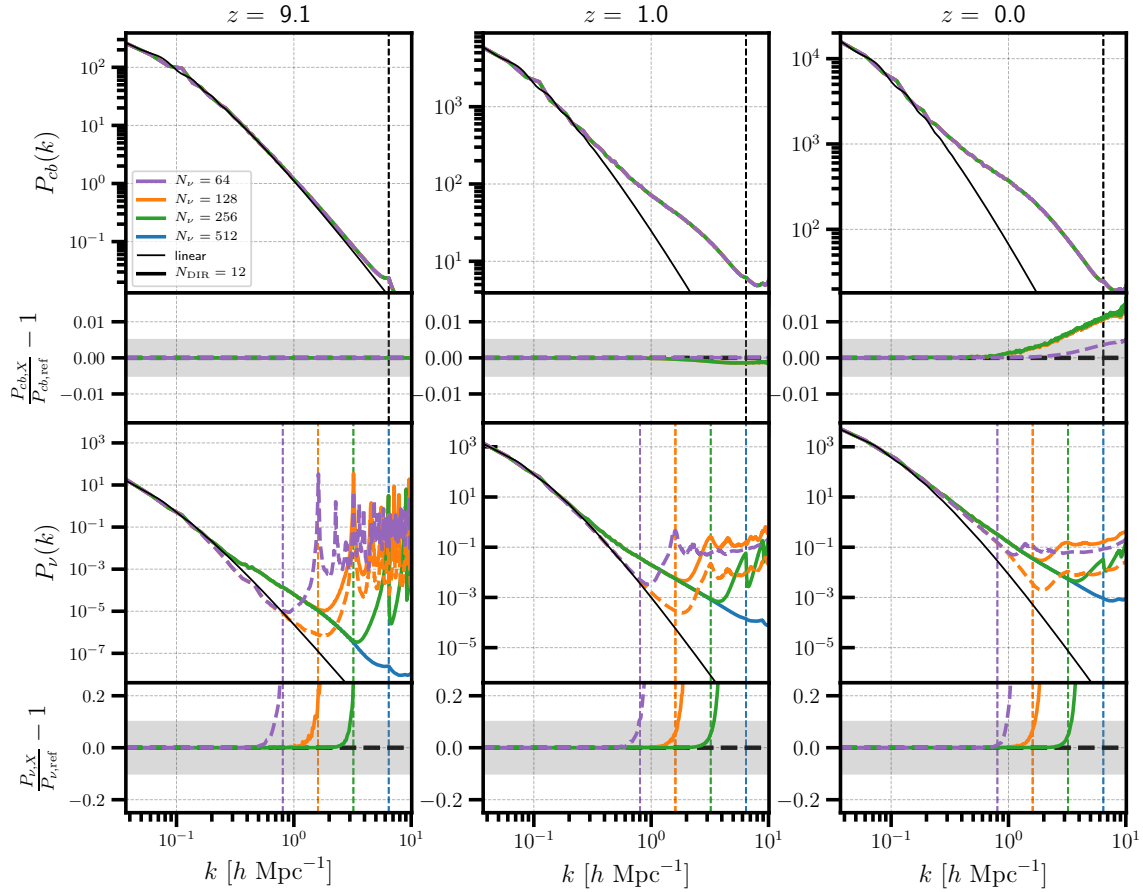
Figure 3.5: Cold matter (upper two rows) and neutrino (lower two rows) power spectra while varying the resolution of the neutrino particle grid, N_ν, for runs with $N_{\mathrm{DIR}} = 12$ (solid lines) and $N_{\mathrm{DIR}} = 192$ (dashed lines). Residuals are shown with respect to the $N_\nu = 512$ run (blue solid line, `hi_Nν`) for the $N_{\mathrm{DIR}} = 12$ runs, and with respect to the $N_\nu = 128$ run (orange dashed line, `coarse_hi_NDIR`) for the $N_{\mathrm{DIR}} = 192$ runs (i.e., they are made with respect to the largest value of N_ν for each set of N_{DIR} runs). Plots are shown at several redshifts in the different columns. Shaded bands in the residual panels are at 0.5% for cold species and 10% for neutrinos. The vertical dashed black line in the upper panels is the Nyquist mode of the initial cold species grid, while in the lower panels, the dotted colored lines correspond to the Nyquist modes of the corresponding neutrino grid resolutions (with colors indicated in the legend). Backscaled linear theory is also shown for reference as the thin black curves.

resolution. We suspect that the probability of the neutrino grid imprinting on the cold matter field is enhanced when the neutrino shot noise is more comparable to the amplitude of the cold matter power spectrum at early times. In other words, care must be taken to control for this issue if the neutrino mass resolution – which, like the shot noise, scales as

$L^3/(N_\nu^3 N_{\mathrm{SHELL}} N_{\mathrm{DIR}})$ – is too coarse.

Of course, the absence of spikes in the cold matter power spectrum of the other two simulations does not mean that the cold matter was not still influenced by the neutrino shot noise level on small scales. This is the motivation behind the use of the sharp-k filter described in Section 3.3 that forcefully removes any contributions from the neutrino grid in the long-range force solver. For reference, the dark red dashed line in Figure 3.6 shows the neutrino power spectrum measured when this filter is applied. The abrupt truncation of the neutrino power spectrum for $k > k_{\mathrm{Nyq}}^\nu$ means that the numerical signal of the neutrino grid is completely hidden from the long-range force calculation. To test the impact of this filter on the final result, the right panels of Figure 3.6 compare P_{cb} and P_ν from the original non-filtered run, `lb_loN_nofil`, with its filtered counterpart, `lb_loN`. Here we observe that the cold matter power spectrum of the filtered run is essentially unchanged below $2k_{\mathrm{Nyq}}$, the location of the first neutrino grid spike. At higher wavenumbers, we see sharp spikes occurring at the Fourier modes of the neutrino grid which correspond to the erasing of these signatures from the cold matter field in the filtered run. At lower redshifts, gravitational broadening of the spikes in the non-filtered run dissipates their signal on small scales. In the lower panel showing the neutrino residuals, we see that the filtering leads to only precent-level changes in the evolution of the neutrino density field. In other words, the filtering strategy effectively removes clear numerical artifacts from the neutrino grid without strongly impacting the small-scale growth of cold matter or neutrino density perturbations.

We reiterate that the sharp-k filter is only justified if neutrinos are not contributing to the growth of density perturbations beyond k_{Nyq}^ν. This assumption becomes increasingly inaccurate the closer k_{Nyq}^ν gets to the neutrino free-streaming scale, and therefore should only be applied if the resolution is such that $k_{\mathrm{Nyq}}^\nu \gg k_{\mathrm{fs}}$. This condition can be somewhat alleviated if the filter is turned off at lower redshift when both the free-streaming length is smaller and the intrinsic neutrino power has grown closer or above the shot noise level. We note that Ref. [215] uses an alternative method to mitigate impacts associated with the neutrino grid which involves radially displacing each neutrino shell during the initialization by an amount proportional to its momentum magnitude. We tried this method but did not find it to be effective for the `lb_loN_nofil` run presented above. In any event, we consider the sharp-k filter to be a more direct method for preventing an artificial back-reaction from the discretization of the neutrino grid. We also tested other functional forms for the filter, but found that the smoothing required to adequately remove the neutrino spikes using obvious candidates (Gaussian and Hann functions) was so broad that the filter artificially suppressed cold matter growth on larger scales.

Emulator Comparison

To make contact between this work and nonlinear growth in total matter power predictions, we compare to the 2022 version of the Mira-Titan Cosmic Emulator [194] in Figure 3.7. The simulations used to train the emulator varied the neutrino overdensity parameter Ω_ν, but were not multi-species simulations, so the influence of neutrinos was incorporated only
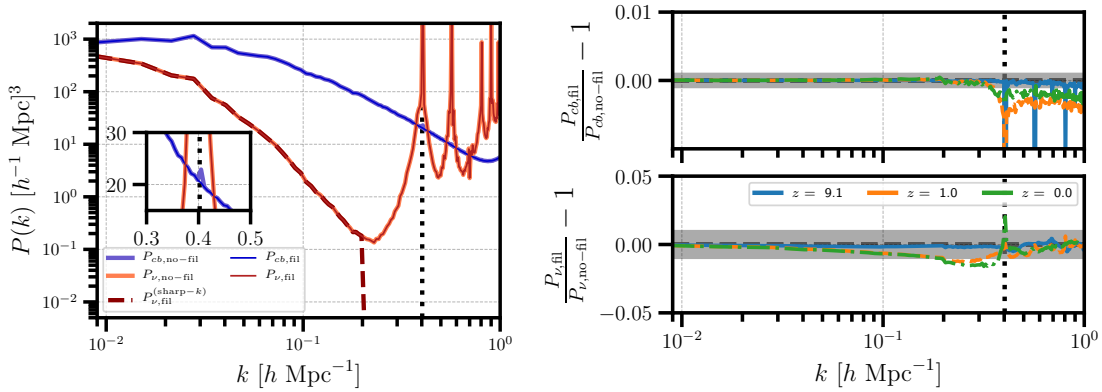
Figure 3.6: *Left*: Cold species (blue) and neutrino (red) power spectra for the `lb_loN_nofil` (thick shaded lines) and `lb_loN` (thin solid lines) runs at $z = 5$. The vertical dotted black line denotes twice the neutrino particle grid Nyquist frequency which corresponds to a prominent spike in the neutrino power spectrum as well as the cold species power spectrum in the non-filtered case (see the inset for a zoom-in on the cold species power spectra). The dashed red line shows the neutrino power spectrum when the sharp-k filter is applied. *Right*: The ratio of the cold species (top) and neutrino (bottom) power spectra from these two runs at redshifts $z = 9.1, 1.0, 0.0$. Shaded regions are at $\pm 0.1\%$ for cold species and $\pm 1\%$ for neutrinos.

through 1) adding the linear massive neutrino density power spectrum to the cold species power spectrum (as described in Refs. [66, 248]), 2) in the computation of σ_8 at $z = 0$, and 3) in the rate of homogeneous and isotropic background expansion. As single-species simulations, these computations neglected scale-dependent growth due to neutrinos in the evolution of the N-body particles, so we do not expect our results for the total matter power to exactly agree with the emulator predictions, aside from the errors intrinsic to the construction of the emulator. Nevertheless, at small enough f_ν, there should be good agreement. Figure 3.7 shows that this is indeed the case at the percent level (here $f_\nu \approx 1\%$ for $\sum m_\nu = 0.15$ eV, where it should be noted that the method used to construct the emulator in the nonlinear regime is effectively an expansion in the leading power of f_ν).

The total matter power spectra for the $N_{\mathrm{SHELL}} = 20, N_{\mathrm{DIR}} = 192$ run at several redshifts approximately match the emulator power spectra at the several percent level. However, the growth in the emulator prediction is systematically higher than the power in our simulation run for wavenumbers that are low enough that the grid does not impact the power spectrum. This offset is at approximately the 5% level, which is larger than the $< 3\%$ accuracy that would be expected from the emulator test set error [194]. This effect is not the result of our implementation of the neutrino N-body particle evolution, as a massless neutrino "noν" cosmology shows a similar level of disagreement in Figure 3.7 (green curve). This disagreement can instead be explained as due to the small box sizes we consider here, as
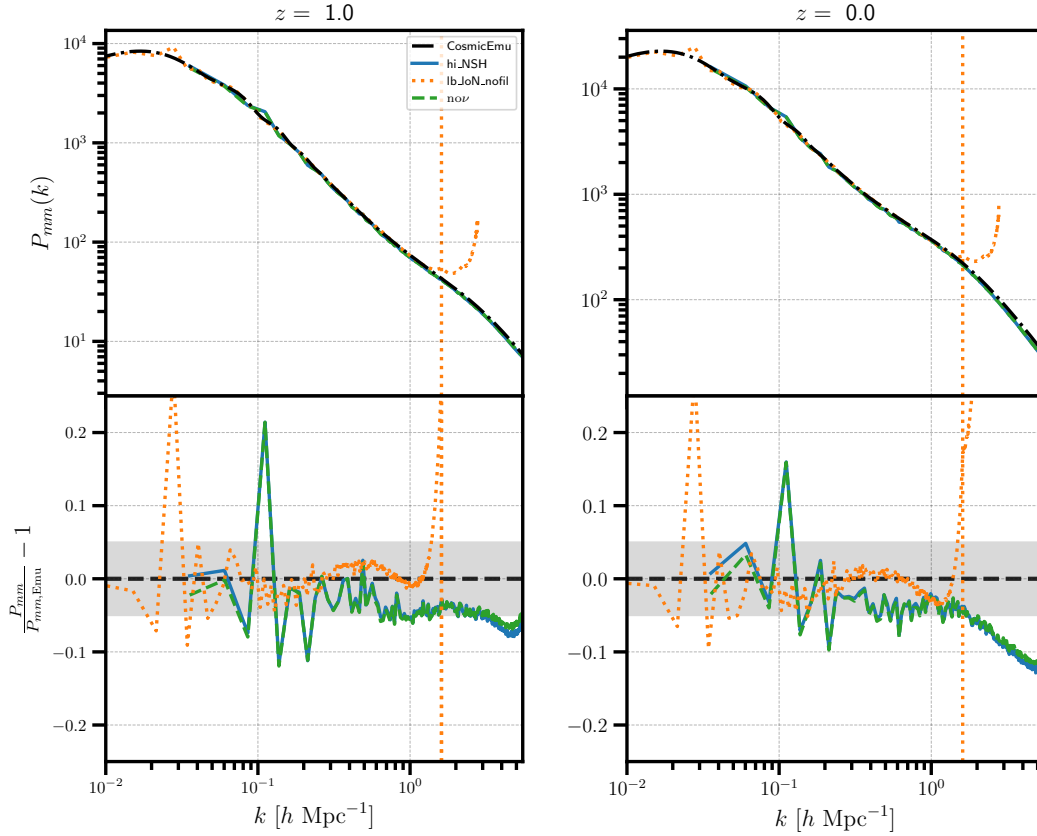
Figure 3.7: Total matter power spectra in the `hi_NSH` ($N_{\mathrm{SHELL}} = 20, N_{\mathrm{DIR}} = 192$) run [blue solid], the `lb_loN_nofil` ($L = 1000\ h^{-1}\mathrm{Mpc}$) run [orange dotted], and the 2022 version of the Mira-Titan CosmicEmu [194] [black dash-dotted] at several redshifts. We also show the matter power spectrum from a neutrino-less cosmology ("noν", with $\Omega_\nu = 0$, $\Omega_{cdm} = 0.27201$) [green dashed] that has the same numerical parameters as the `fid_nofil` run. The orange vertical line shows the Nyquist wavenumber for the `lb_loN_nofil` run. Lower panels show residuals with shaded bands between $\pm 5\%$. The residual for the "noν" line is with respect to the CosmicEmu prediction for a neutrino-less cosmology (not shown in the upper panel). See the text for a discussion of the finite-volume effects seen here (the shift between the blue and orange curves).

in one of our larger box simulations (orange dotted curve), we see no such 5% offset, and any disagreement between the simulation run and the emulator is at the expected level of $< 3\%$ (below the grid scale of that simulation, which is $k \sim 1.0\ h\ \mathrm{Mpc}^{-1}$). These types of finite-volume effects are characterized in more detail in Section 3.3 of Ref. [249], and can be understood more broadly as a combination of a super-sample mode effect [250, 251, 252, 83, 253], as the small-box simulation is effectively missing large-scale power that is present in

the larger simulation box (see e.g. Fig. 6 of Ref. [252]) and increased variance in individual realizations at smaller box sizes. The estimated variation from the results of Ref. [249] is consistent with that observed in the comparison. In conclusion, Fig. 3.7 shows that the emulator results are consistent with the direct simulations as performed here within the level of errors that are characteristic of both methods.

Examination of Neutrino Growth

We complete this section with a closer look at how well the simulated neutrino growth compares to linear theory expectations. In Figure 3.8, we compare the linear neutrino power spectra at various redshifts to the outputs from the `fid`, `med_NDIR`, and `hi_NDIR` runs (previously shown in Figure 3.3). As discussed earlier, the neutrino power spectrum on scales approaching k^{ν}_{Nyq} is very sensitive to the number of direction vectors. Ref. [213] attributed this to a coherent sampling of the gravitational potential causing a spurious power generation that diminishes as the number of direction vectors increases.[6] At high redshift, we would expect that the simulated power spectra numerically converge to linear theory as the number of direction vectors increases, and that this happens most readily on larger scales where the spurious power is weakest. However, while we do observe clear numerical convergence as N_{DIR} is increased, the converged result is noticeably *below* linear theory for $k \gtrsim 0.1~h~\mathrm{Mpc}^{-1}$. This systematic suppression is clear at $z = 9$ and persists until late times when non-linear growth eventually dominates the signal. While this discrepancy is relatively minor (the simulation with $N_{\mathrm{DIR}} = 192$ has neutrino power that is 15% smaller than linear theory), and therefore unlikely to significantly impact the total matter field, it is worth considering any potential systematic errors that could contribute to incorrect neutrino growth.

There are a number of approximations made in the simulations presented here. In the first place, the momentum-integrated neutrino transfer function is used to set the initial displacement and non-thermal velocities for all momentum shells. In principle, this could be made more accurate by initializing each shell with its own momentum-dependent transfer function. However, given that we do not observe a significant dependence on N_{SHELL} seen in Figure 3.4, we do not expect this to be a large source of bias in the neutrino power spectrum. Another potential issue is that high thermal motion at early times enables neutrino particles to quickly traverse large distances from their starting point. This behavior may smooth out the power spectrum initially seeded in the displacement field as the neutrinos dynamically readjust to their rapidly changing environments. In this case, we would expect the smoothing to be enhanced with earlier starting times when the thermal motion is more extreme. Another possible bias stems from the non-relativistic treatment of neutrino particles in the simulation; an approximation that becomes increasingly inaccurate at higher redshift and conflicts with the relativistic treatment used in the linear theory reference curves. Finally, the neutrino

---

[6]Ref. [213] also found that this spurious power is more significant for higher momentum shells. By inspecting $q-$dependent power spectra, we also find this to be the case, and that the higher-q shells are more sensitive to the number of directions N_{DIR}.

initial conditions are set using the ZA which is known to be insufficient for cold matter at the neutrino starting redshifts used here. It is unclear if the ZA is also problematic for neutrinos at these redshifts despite their lower density amplitude, but it has been shown in Ref. [244] that systematic power suppression occurs in cold matter simulations with late ZA starts.

We test the latter three possibilities more closely by running three simulations with different neutrino initial redshifts. These are shown in the right panel of Figure 3.8 where the `lo_z`, `hi_NDIR`, and `hi_z` runs differ only by their use of $z_{\text{ini}}^{\nu} = 10$, 20, and 40, respectively. At $z = 9$, all three runs sit systematically below linear theory[7] at $k \sim 0.2\ h\ \text{Mpc}^{-1}$, with this suppression persisting until non-linear growth dominates at $z = 1$. We do observe a small dependence on z_{ini}^{ν} at high redshift, with the `lo_z` run exhibiting tighter agreement with linear theory compared to the other two simulations. However, this trend is quickly erased by $z = 3$ at which point all three runs show strong agreement on all but the smallest scales. On scales approaching k_{Nyq}^{ν}, we still see a clear trend with z_{ini}^{ν} which we attribute to the earlier start times allowing a greater accumulation of spurious power due to the angular discretization of the tiling method. Hence, the systematic suppression in power does show minor dependence on the neutrino initial redshift, but this is only evident at early times.

More careful tests would be required to determine the exact cause of the neutrino power suppression seen on intermediate scales. Furthermore, it is not clear if this suppression is specific to the tiling scheme or would also be observed in a simulation using the random draw strategy (assuming a sufficient number of particles are used to resolve the neutrino power below the shot noise level). We note that analogous findings have also been observed in mixed cold dark matter plus baryon simulations where the cold dark matter (baryons) grow systematically fast (slow) relative to linear theory expectations when the force resolution is below the mean inter-particle separation [254, 255, 256]. It is conceivable that the same issue manifests in neutrino simulations and may even be exacerbated when the neutrino particle grid is made coarser than that of the cold matter. Even though the discrepancy with respect to linear theory is relatively minor, it is still worth investigating this topic in the future so that systematics in neutrino simulations are properly identified and resolved.

3.5   Conclusions

A standard approach for initializing cosmological simulations is to backscale the final redshift transfer functions to the initial redshift in a manner that is consistent with the Newtonian forward model of the simulation. This procedure becomes complicated in massive neutrino cosmologies since the simulated growth function acquires a scale dependence that is not easily calculable. We have presented an improved backscaling method for massive neutrino cosmologies that uses an iterative procedure to converge on a self-consistent initialization strategy that preserves the relative neutrino growth predicted by Boltzmann solvers. Our

---

[7]The linear theory curves are calculated from the backscaling of the $z_{\text{ini}}^{\nu} = 20$ run. There are minor differences in the backscaling for the different z_{ini}^{ν} (see Figure 3.1) but these are small in comparison to the spread in the simulation curves seen here.
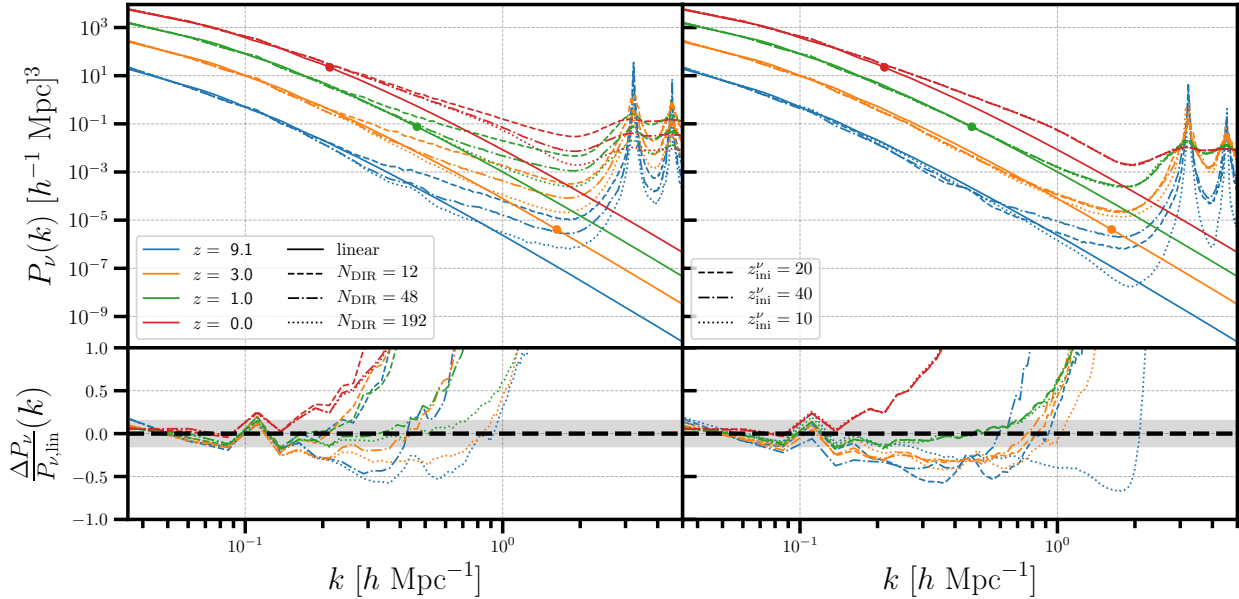
Figure 3.8: Neutrino power spectra from several simulation runs at several redshifts (colors) compared to linear theory computed using our backscaled transfer functions (solid). *Left:* Neutrino power for the fiducial neutrino initialization redshift ($z_{\mathrm{ini}}^{\nu} = 20$), with a varying number of initial velocity directions: $N_{\mathrm{DIR}} = 12$ (dashed, `fid`), $N_{\mathrm{DIR}} = 48$ (dash-dotted, `mid_NDIR`), and $N_{\mathrm{DIR}} = 192$ (dotted, `hi_NDIR`). *Right:* Neutrino power for the highest number of initial velocity directions ($N_{\mathrm{DIR}} = 192$) at several choices of starting redshift for the neutrino particles: $z_{\mathrm{ini}}^{\nu} = 20$ (dashed, `hi_NDIR`), $z_{\mathrm{ini}}^{\nu} = 40$ (dash-dotted,, `hi_z`), $z_{\mathrm{ini}}^{\nu} = 10$ (dotted,`lo_z`). At each redshift, we show using a circle the non-linear scale, k_{nl}, defined to be the largest scale for which the dimensionless total matter power spectrum, $\Delta_m^2(z) \equiv k^3 P_m(k)/2\pi^2$, exceeds unity. Lower panels show residuals with respect to (backscaled) linear theory neutrino power spectra and gray bands are between $\pm15\%$.

method exploits the known asymptotic limits that the energy density contributing to growth on large and small scales is set by $\Omega_{cb\nu}$ and Ω_{cb}, respectively. When compared to the direct output of Boltzmann solvers at high redshift, our backscaled transfer functions show qualitatively similar behavior to what is seen in massless neutrino cosmologies. This is a notable improvement compared to the common approach where the scale-dependent growth is solved using a two-fluid model that assumes a much different level of neutrino growth compared to Boltzmann solvers. Since our approach maintains the neutrino growth predicted with Boltzmann codes, an added benefit is that we can readily compare the simulated neutrino power spectra with our backscaling model to check that neutrinos are evolving with linear theory expectations.

We applied this backscaling procedure to particle-based neutrino simulations performed using the code HACC. The neutrino initial conditions follow the tiling framework presented

in Ref. [224] with a number of important modifications. First, we introduce a simple rotation
scheme between adjacent momentum shells which reduces spurious power arising from the
angular discretization of momentum directions. Second, we omit neutrinos from the short-
range gravitational force and apply a sharp truncation to the neutrino density field in Fourier
space during the long-range force calculation. This filter can easily be applied in any PM-
based force solver and prevents an artificial back-reaction of the initial condition neutrino
grid onto the cold matter field. Without any filter, we found perceptible spikes emerge in
the cold matter field if the neutrino mass resolution, and the corresponding neutrino shot
noise level, is sufficiently coarse. Since this filter completely removes neutrino power below
the particle grid Nyquist frequency, it should only be applied if the particle grid resolution
is much finer than the neutrino free-streaming length. More sophisticated treatments would
involve applying this filter only at high redshift when both the free-streaming length is
larger and the neutrino shot noise level is more comparable to the intrinsic cold matter
power spectrum.

We performed a numerical convergence study on the internal parameters of the tiling
method. In agreement with Ref. [213], we find that the tiling method is highly sensitive to
the number of direction vectors and $N_{\mathrm{DIR}} \gtrsim 100$ is required to achieve $\sim 10\%$ convergence
in the neutrino power spectrum on intermediate scales. The dependence on the number of
momentum shells is weaker and is generally converged at the $\sim 10\%$ level with $N_{\mathrm{SHELL}} \gtrsim 10$.
In order to prevent the total number of neutrino particles from heavily dominating the total
particle count, these findings require a relatively coarse neutrino grid with $N_{cb}/N_\nu \gtrsim 10$.
The primary challenge is that the neutrino power spectrum saturates to the shot noise level
below the neutrino grid scale with prominent peaks at the grid resonance frequencies. This
issue is not unique to the tiling procedure per se, but becomes increasingly problematic as
a neutrino grid pushes the noise to larger scales. Hence, while the tiling method effectively
removes neutrino shot noise above the neutrino grid scale, the sacrifice is that care must be
taken to ensure that the number of neutrino particles per grid site is large enough to reach
numerical convergence and that the gravitational back-reaction from the grid is properly
handled. We found that the cold matter power spectrum is largely invariant to changes in
the tiling parameters on linear scales. On non-linear scales, however, we found that changes
in the tiling parameter choices lead to $\sim 1\%$ random noise.

The simulations presented here converge to a neutrino power spectrum suppressed at the
15% level compared to linear theory for scales $k \gtrsim 0.1\ h\ \mathrm{Mpc}^{-1}$ and redshifts $z \gtrsim 1$. We
have not determined the exact cause of this discrepancy nor is it clear if this result is unique
to the tiling method or would more generally be reproduced in other particle-based methods.
Obviously, there are a number of approximations made in the simulation that could preclude
accurate comparisons with linear theory. These include the use of the momentum-integrated
transfer function for all momentum bins in the initial conditions as well as the fact that we
omit all relativistic treatments of neutrinos. It is also possible that high thermal motion
at early times erases the initial condition power spectrum and/or that biases seen in mixed
dark matter plus baryon simulations extend to the neutrino case. The use of the ZA in
initializing the neutrinos at $z_{\mathrm{ini}}^\nu = 10 - 40$ may also be problematic, due to the associated

power suppression characteristic of this approximation. Understandably, most studies have focused only on numerical convergence in the total matter field, but it remains important to isolate any errors directly impacting neutrino growth.

We compared the results from our N-body simulations for the total matter power spectrum with the predictions of the Mira-Titan emulator [194] for a neutrino mass sum of 0.15 eV, corresponding to $f_\nu \approx 1\%$. The results obtained are very encouraging, the agreement being nicely within the estimated errors intrinsic to both methods.

Looking forward, there are several applications where the lessons of this work might be useful. For example, the methods used for simulating the nonlinear effect of massive neutrinos on probes of LSS can be repurposed to investigate similar effects due to light massive relics (LiMRs, e.g. Ref. [257]). Most forecasts of the sensitivity of LSS observations to the presence of LiMRs have been at the level of linear theory or modified perturbation theory [258, 259], though Ref. [260] recently gave a fully-nonlinear treatment with simulations (using a randomized initialization). It would be interesting to consider both the reduced computational cost due to shot noise reduction and to explore the relative significance of the numerical artifacts of the tiling initialization in the context of LiMR simulations. This work also serves to inform how best to include massive neutrinos in a computationally efficient manner in three-species simulations, for example, in conjunction with CRK-HACC [239]. Future work will more completely characterize to what extent the tiling initialization can be reliably applied to three-species simulations, in particular focusing on the interaction between baryons and neutrinos.

Acknowledgments

Chapter 4

Multi-tracer Forecasts for Primordial non-Gaussianity with Machine-Learned Bias

This chapter was originally published as: James M. Sullivan, Tijan Prijon, and Uroš Seljak. "Learning to Concentrate: Multi-tracer Forecasts on Local Primordial Non-Gaussianity with Machine-Learned Bias". In JCAP 2023, 8, 004, DOI:10.1088/1475-7516/2023/08/004, arXiv:2303.08901 [astro-ph.CO]

In this chapter, we discuss the prospect of using the galaxy properties beyond host halo mass of multiple tracers of large scale structure for constraining inflation-era physics. Local primordial non-Gaussianity (LPNG) is predicted by many non-minimal models of inflation, and creates a scale-dependent contribution to the power spectrum of large-scale structure (LSS) tracers, whose amplitude is characterized by b_ϕ. Knowledge of b_ϕ for the observed tracer population is therefore crucial for learning about inflation from LSS. Recently, it has been shown that the relationship between linear bias b_1 and b_ϕ for simulated halos exhibits significant secondary dependence on halo concentration. We leverage this fact to forecast multi-tracer constraints on $f_{NL}^{\rm loc}$. We train a machine learning model on observable properties of simulated IllustrisTNG galaxies to predict b_ϕ for samples constructed to approximate DESI emission line galaxies (ELGs) and luminous red galaxies (LRGs). We find $\sigma(f_{NL}^{\rm loc}) = 2.3$, and $\sigma(f_{NL}^{\rm loc}) = 3.7$, respectively. These forecasted errors are roughly factors of 3, and 35% improvements over the single-tracer case for each sample, respectively. When considering both ELGs and LRGs in their overlap region, we forecast $\sigma(f_{NL}^{\rm loc}) = 1.5$ is attainable with our learned model, more than a factor of 3 improvement over the single-tracer case, while the ideal split by b_ϕ could reach $\sigma(f_{NL}^{\rm loc}) < 1$. We also perform multi-tracer forecasts for upcoming spectroscopic surveys targeting LPNG (MegaMapper, SPHEREx) and show that splitting tracer samples by b_ϕ can lead to an order-of-magnitude reduction in projected $\sigma(f_{NL}^{\rm loc})$ for these surveys.

4.1 Introduction

Cosmological observations of the Cosmic Microwave Background (CMB) and Large-scale Structure (LSS) are currently in complete concordance with Gaussian initial conditions [261]. Yet, the tantalizing possibility that there are deviations from Gaussianity lurking beneath observational uncertainties remains. These deviations - or Primordial non-Gaussianity (PNG) - come in several well-known forms, one of which is the *local* type (LPNG) parameterized by the amplitude $f_{NL}^{\rm loc}$:

$$\phi = \phi_G + f_{NL}^{\rm loc} \left[\phi_G^2 - \langle \phi_G^2 \rangle \right], \tag{4.1}$$

where ϕ is the Bardeen potential [262] and ϕ_G a Gaussian random field. A detection of $f_{NL}^{\rm loc} \sim 1$ would definitively indicate the presence of multiple fields during inflation, and provides a natural theoretical sensitivity target [263, 264].

The Planck satellite has placed the most stringent constraints on $f_{NL}^{\rm loc}$ to date, finding $f_{NL}^{\rm loc} = -0.9 \pm 5.1$ [265], consistent with no local PNG. While future CMB missions should improve on this uncertainty, they are not forecasted to reach $\sigma(f_{NL}^{\rm loc}) \lesssim 1$ [266, 267]. However, LSS surveys promise to supersede this sensitivity in the near-term [268, 269, 270, 271, 272, 273, 274]. In particular, the multi-tracer technique [275, 276, 277, 278, 274, 273, 279, 280, 281, 282] provides a powerful tool for using large-scale modes most affected by survey sample variance.

The presence of local PNG modulates the halo density field in a scale-dependent manner [283, 284]. This "scale-dependent bias" effect (LPNG bias) produces a signal in the halo power spectrum that scales like k^{-2}, becoming more important on larger scales. This behavior is fundamentally due to LPNG inducing a bispectrum that peaks in the squeezed triangle configuration, which couples a long wavelength mode with two short wavelength modes, effectively linking very large scales with the small-scale process of halo formation. The amplitude of this effect is controlled by the LPNG bias b_ϕ, and while initially it was believed that it only depends on the halo mass [283], it was soon realized secondary halo properties such as merger history also affect its value [284]. This effect has been exploited over the last 15 years in several LSS analyses to make significant progress toward a precision measurement of $f_{NL}^{\rm loc}$ [284, 285, 286, 287, 288, 289], though those constraints remain looser than those from Planck - LSS has yet to take the lead in constraining $f_{NL}^{\rm loc}$. The above analyses do not make use of the multi-tracer technique, and therefore can in principle be pushed further in sensitivity. However, many these analyses make a simplifying assumption that the relationship between the LPNG bias b_ϕ and the linear tracer bias b is perfectly known, or otherwise only consider a few values of constant shifts from this relationship [1].

A slew of recent papers has called this knowledge, which was always approximate, into significant question for simulated galaxies and other tracers [291, 292, 293, 290, 294]. Building on the pioneering work of Ref. [295], Ref. [296] investigated to what extent several commonly considered assembly bias parameters affect the relationship between the linear halo

---

[1]Though more recent work has also considered priors on b_ϕ and constraining only $b_\phi f_{NL}^{\rm loc}$ [290, 289, 288] to account for this uncertainty

bias b_1 and b_ϕ in dark matter only N-body simulations. For halo spin and sphericity, the authors found only a modest effect, but, especially for lower halo masses and redshifts, they found a significant secondary dependence of b_ϕ on halo concentration. There, the authors of Ref. [296] argue that the strong sensitivity of b_ϕ to changes in concentration indicates that existing constraints assuming a fixed form for $b_\phi(b_1)$ are unreliable, but also acknowledge that this sensitivity is an opportunity to better constrain f_{NL}^{loc} with suitably defined galaxy populations. The goal of this work is to identify such populations with the aid of a machine learning algorithm trained on observable galaxy properties and to use its predictions to forecast the improvement on $\sigma(f_{NL}^{\text{loc}})$ that such a definition furnishes.

By using simulated galaxy samples that approximate the DESI selections for emission line galaxies (ELGs) and luminous red galaxies (LRGs), we can get a realistic estimate for the current and near-term prospects of constraining f_{NL}^{loc} by using observable quantities beyond halo mass to which b_ϕ responds. We will show that, depending on the information available and the galaxy sample properties, we are able to improve the forecasted error on f_{NL}^{loc} by factors of several for DESI-like galaxies using machine-learned b_ϕ predictions. We also comment on the prospects for using this strategy in future spectroscopic surveys, where naively one could expect an order of magnitude improvement on $\sigma(f_{NL}^{\text{loc}})$ if halo concentration were perfectly recoverable.

This paper is structured as follows: We briefly review the LPNG parameter b_ϕ in Section 4.2 before moving on to the construction of the DESI-like simulated galaxy samples and our machine learning methodology in Section 4.3. We then provide forecasts for $\sigma(f_{NL}^{\text{loc}})$ in Section 4.4, and provide some concluding remarks in Section 4.5.

4.2 Local PNG bias b_ϕ

We briefly review the salient facts about the scale-dependent bias signal induced by local primordial non-Gaussianity and its amplitude b_ϕ.

In the peak-background split (PBS) formalism [21], the parameter b_ϕ is defined as the response of the tracer mean density $\bar{n}_t$ to the presence of a long-wavelength perturbation of the Bardeen potential ϕ

$$b_\phi = \frac{\partial \log \bar{n}_t}{\partial (f_{NL}^{\text{loc}} \phi)} = 2 \frac{\partial \log \bar{n}_t}{\partial \log \sigma_8}. \tag{4.2}$$

The second equality follows from the argument that the effect of LPNG on the tracer abundance is equivalent to the effect of a rescaling of the amplitude of linear fluctuations [21]. We will accept this argument for the purposes of this work.

This bias enters the tracer power spectrum at linear order in δ and f_{NL}^{loc}:

$$P_{tt}(k) = \langle \delta_t(\mathbf{k}) \delta_t^*(\mathbf{k}') \rangle' \tag{4.3}$$

$$\overset{\text{L.O.}}{=} b^2 P_L(k) + 2 b_\phi b \mathcal{M}^{-1}(k) P_L(k), \tag{4.4}$$

where the $'$ indicates that we drop the missing momentum-conserving Dirac delta and associated $(2\pi)^3$ prefactor, b is the linear bias, and $\mathcal{M}^{-1}(k) = \frac{3\Omega_m H_0^2}{2c^2 D_{md} k^2 \tilde{T}(k)}$, where $\tilde{T}(k)$ is the

transfer function normalized to 1 on large scales, and D_{md} is the linear total matter growth factor normalized to 1 in matter domination.

Reducing to the case of halos ($\bar{n}_t = \bar{n}_h$) with a universal mass function that only allows halo abundance to depend on halo mass, gives the universality relation (or UMF [284]):

$$b_\phi(b, p) = 2\delta_c(b - p),$$

where the critical overdensity is $\delta_c = 1.686$. For the case where we select all of the halos of a given mass the universality relation gives $p = 1$. For dark matter halos there are some deviations of this $b_\phi(b_1)$ relationship observed [297, 290]. It has also been argued that other values of p are more appropriate for certain sub-populations of halos, such as for halos that have undergone a recent merger, which may host quasars [284]. However, the picture becomes even more complicated for simulated galaxies, for which it was recently shown that when selecting by stellar mass, a value of $p = 0.55$ is preferred when using the IllustrisTNG galaxy formation model [291]. Presumably, if the stellar mass selection can be replaced with halo mass selection, this would restore $p = 1$, which suggests that by choosing additional observables one may be able to split galaxies by b_ϕ.

One such additional observable that has been found to be very sensitive to b_ϕ is halo concentration c. With a fixed galaxy formation model, a value of the LPNG bias can be estimated from Separate Universe (SU) simulations [e.g., 298, 299, 300, 301, 302, 303, 304, 305] in which two N-body simulations are run with two values of σ_8 and the bias is then estimated via finite difference with respect to the mean number density $\bar{n}_h(M, c)$. We use the reported results of Refs. [296, 291] at $z = 1$ to relate linear halo bias b_1 (estimated from power spectra) and SU LPNG bias b_ϕ to halo mass M and concentration c.

4.3 b_ϕ with machine learning

In this Section, we describe our machine learning method for obtaining predictions for b_ϕ. Before diving into the details, we provide a brief overview of this process here.

First, we obtain $b(M, c), b_\phi(M, c)$ by interpolating the gravity-only Separate Universe (SU) results of [296] in halo mass M and concentration c, and assign $b(M, c), b_\phi(M, c)$ to each TNG galaxy's host halo of mass M and concentration c . We then train the ML model described below on galaxies where the values of b_ϕ are known to learn the relationship between input features such as galaxy magnitude and b_ϕ. After making a prediction for b_ϕ on test data, we rank order the b_ϕ predictions to determine a split of the sample into tertiles. We then obtain predictions for a mean value of b_ϕ in each tertile by using the trained ML model evaluated on the observable input features of galaxies in that tertile (magnitude, position, etc.) and averaging. This final set of b_ϕ values in the tertiles are used as input to the Fisher forecasts of Section 4.4.

Simulated galaxy sample

Simulation
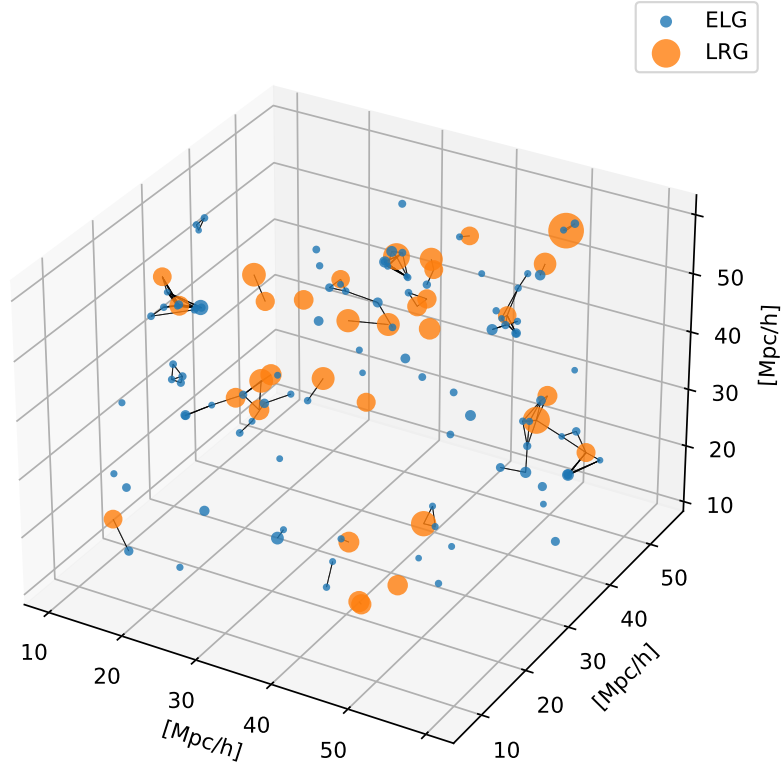
Figure 4.1: ELG and LRG-like galaxies in the part of the volume of the IllustrisTNG simulations at redshift $z = 1$. Galaxies are presented in real space, and those closer than $5\ h^{-1}$ Mpc are linked for ease of presentation. The size of each galaxy marker is directly proportional to its logarithmic stellar mass ($\log M^*$).

We use simulated galaxy data from the IllustrisTNG-300 [306, 307, 308, 309, 310, 311] simulation, a cosmological magnetohydrodynamical simulation that models the formation and evolution of galaxies and covers the volume of $V = 205^3\ [h^{-1}\ \mathrm{Mpc}]^3$. The snapshot that we use is at redshift $z = 1$, which is a compromise between being able to mimic Dark Energy Spectroscopic Instrument (DESI) survey [312] observations of both emission-line galaxies (ELGs) and luminous red galaxies (LRGs). A sub-volume of the simulation containing ELG and LRG- like galaxies, as we define in Section 4.3, is shown in Fig. 4.1.

Observables

The halos in the IllustrisTNG simulation are determined with the friend-of-friends (FoF) algorithm [313], with linking length $\ell = 0.2$.

We use the results of Ref. [296] to link b_ϕ and b to halo mass and concentration. We take halo concentration from the IllustrisTNG supplementary data catalog [314]. It is defined as $c = R_{200c}/R_s$, where R_s is determined by fitting a Navarro-Frenk-White (NFW) [315] profile to the dark matter density profile. We use the R_{200c} halo mass definition. We note that Ref. [296] used several halo mass definitions when using different simulations, though since concentration was logarithmized, the effect of these differences should be reduced.

To avoid using halo properties that are unobservable by spectroscopic surveys to infer halo LPNG bias, we only use stellar masses M^*, the r, g, and z magnitudes, and redshift-space positions of the galaxies. We use the word "observable" somewhat loosely, as galaxy stellar mass must be estimated from spectra, and dust models that affect the observed r, g, z magnitudes are uncertain - we do not account for these aspects that would be relevant for a real data analysis here. We choose the IllustrisTNG z-coordinate as the line-of-sight direction to obtain positions in redshift space. The real-to-redshift space transformation is

$$\mathbf{s} = \mathbf{x} + \frac{v(\mathbf{x})\mu}{\mathcal{H}}\hat{z} \tag{4.5}$$

where $\mathbf{x}$ is the real-space galaxy position, $v(x)$ is the galaxy velocity magnitude, $\hat{z}$ is the line-of-sight direction, $\mu = \hat{\mathbf{v}} \cdot \hat{z}$, $\mathcal{H}$ is the conformal Hubble parameter, and $\mathbf{s}$ is the redshift-space position. To simulate the effect of redshift-space distortions on galaxy observations, we transform the galaxies to redshift space before computing input features and training the machine learning algorithm to predict b_ϕ.

While each galaxy lies in exactly one halo, a halo can contain multiple galaxies inside. When we report results in later sections, we predict b_ϕ of the halo for each galaxy separately. Thus, it can happen that in the results, we have several (either the same or different) predictions for the same halo if several galaxies lie in it. This approach is justified because we, in reality, observe galaxies and cannot a priori know whether two galaxies belong to the same halo. We emphasize that in the training we do not use any information about the halos (for example, halo position), except the knowledge of to which parent halo a galaxy belongs.

ELG and LRG sample

Our data sample mimics the emission-line galaxies (ELGs), and luminous red galaxies (LRGs), as observed by the DESI survey at redshift $z = 1$. We can construct these mock galaxy samples either by color-space cuts or by galaxy selection based on specific star-formation rate (sSFR, i.e., the SFR per stellar mass) and stellar mass (M^*) cuts [316, 317, 318]. We adopt the latter strategy and use sSFR and stellar mass cuts.

To obtain both ELG and LRG samples, we first make sSFR cuts and then take the heaviest galaxies based on M^* to get the desired target number densities. We made two
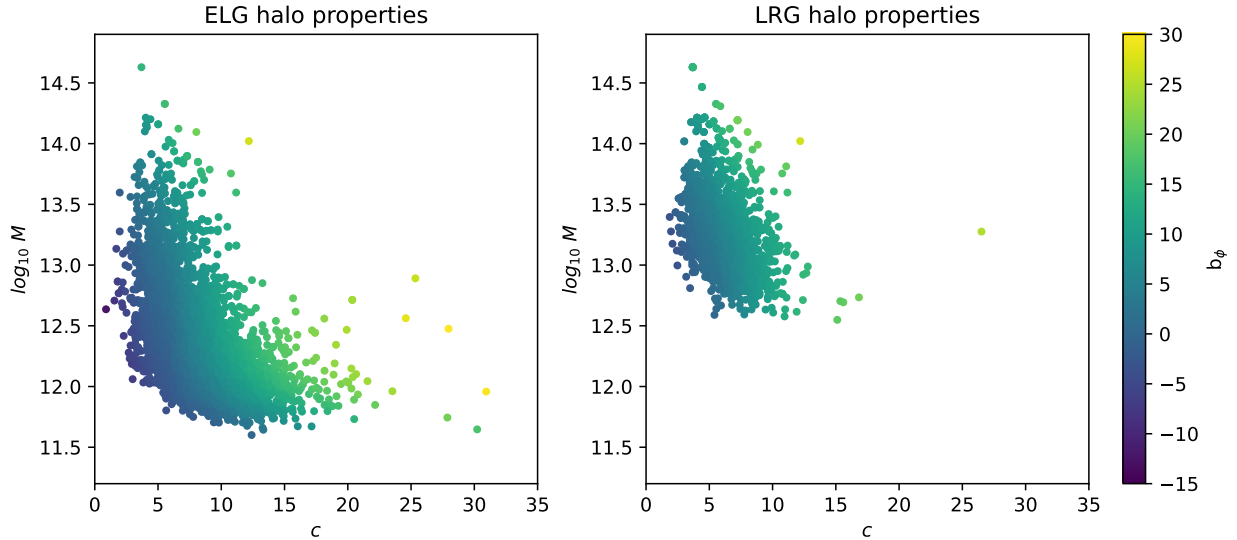
Figure 4.2: Parent halo properties of ELGs and LRGs. As expected we see that LRGs live in more massive halos with lower concentration. We also show the b_ϕ value associated each halo. We can see that while it is a function of both halo mass and concentration, the latter has a much greater impact. The halo mass is reported in the units of 10^{10} M$_\odot/h$.

different sample selections using sSFR cut values of $\log_{10}[sSFR] = -9.09$ from [318], and $\log_{10}[sSFR] = -9.23$ from [316]. The target number densities are taken from [319, 320]. The selection of ELG and LRG samples is summarized in Table 4.1. Here we find that the mean b_1 values of our mock samples are similar to those of Ref. [312] for ELGs (1.4) and LRGs (2.6) at $z = 1$. With this selection, we also ensure that no galaxy is classified as an ELG and an LRG simultaneously. In the next sections we report the results for all eight galaxy samples presented here. We optimize hyperparameter separately for each sample (see Section 4.3). However, because the samples are relatively similar, most hyperparameters remain the same.

In Fig. 4.2, we compare the halo properties in which LRG and ELG galaxies lie, and show their expected b_ϕ in the mass and concentration plane. We see an apparent trend that LRGs lie in heavier halos and that their halos have lower concentrations. In Fig. 4.3, we compare the stellar mass of ELGs and LRGs and its difference when applying different sSFR cuts. While the difference in sSFR affects only the M^* distribution of ELGs, it is evident that LRGs are significantly heavier, in accordance with Refs. [312, 321].

All the results in this Section are reported for the samples selected with the sSFR cut of $\log_{10} sSFR = -9.09$, and with number densities of $n_{ELG} = 5 \times 10^{-4}$ $[h^{-1}$ Mpc$]^{-3}$, and $n_{LRG} = 2 \times 10^{-4}$ $[h^{-1}$ Mpc$]^{-3}$, as these are the DESI number densities at $z = 1$ in Refs. [320, 319].

Type	$\log_{10} sSFR$	$n \ [h^{-1} \ \mathrm{Mpc}]^{-3}$	$\min \log_{10} M^*$	$\bar{b}$
LRG	$<$ -9.09	2×10^{-4}	1.06	2.40
LRG	$<$ -9.23	2×10^{-4}	1.06	2.40
ELG	$>$ -9.09	5×10^{-4}	0.02	1.44
ELG	$>$ -9.09	7×10^{-4}	-0.09	1.42
ELG	$>$ -9.09	1×10^{-3}	-0.23	1.39
ELG	$>$ -9.23	5×10^{-4}	0.23	1.46
ELG	$>$ -9.23	7×10^{-4}	0.17	1.44
ELG	$>$ -9.23	1×10^{-3}	0.07	1.42

Table 4.1: Selection criteria with number densities, minimum obtained stellar mass M^* (units of $10^{10} \ M_\odot h$), and average b_1 for each sample. We first apply the sSFR cut (units of h/yr), as specified in the second column, and then select galaxies with the highest M^* to achieve the target number density. The number of galaxies ("size of dataset") in each sample can be calculated from number densities and the volume of the simulation ($V = 205^3 \ [h^{-1} \ \mathrm{Mpc}]^3$), and lie between 1500 and 7000.

Machine learning

Our objective is to predict the value of the continuous variable b_ϕ, based on M continuous features (inputs) $\mathbf{x_i} = (x_{i,1}, x_{i,2}, ..., x_{i,M})$ on N training examples, which are in our case galaxies. We want to find a function $f : \mathbb{R}^M \to \mathbb{R}$ that maps the inputs $\mathbf{x_i}$ to the scalar value of b_ϕ. When constructing such a function, we want to minimize the difference between its output and the actual value of b_ϕ. The metric of the difference we are minimizing is our loss function, which we take to be the most common root-mean-square error function (RMSE), defined as $L = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2}$, where y_i and $\hat{y}_i$ are the actual and predicted values of b_ϕ for the ith training example.

The described problem is a single-target regression problem. However, rather than first interpolating b_ϕ from halo mass and concentration (as described in Section 4.3) and directly predicting it, we could make predictions for mass and concentration, and interpolate b_ϕ from predicted values instead. In this case, we would be dealing with a multi-target regression problem. We tried this approach and verified that it leads to worse results according to the RMSE metric. The difference between the two approaches can be imagined only as a difference in the loss function that is optimized. However, because we are only interested in b_ϕ it is both simpler and more effective to focus on a single target regression.

The problem of inferring the halo properties from observable data has already been studied in the literature using various machine learning (ML) techniques. One approach is to use structured data with a combination of graph neural networks (GNN) or convolutional neural networks (CNN), while another approach uses unstructured data by calculating statistics of the halo neighborhood [e.g., 322, 323, 324, 325, 326, 327, 328]. We focus on the latter
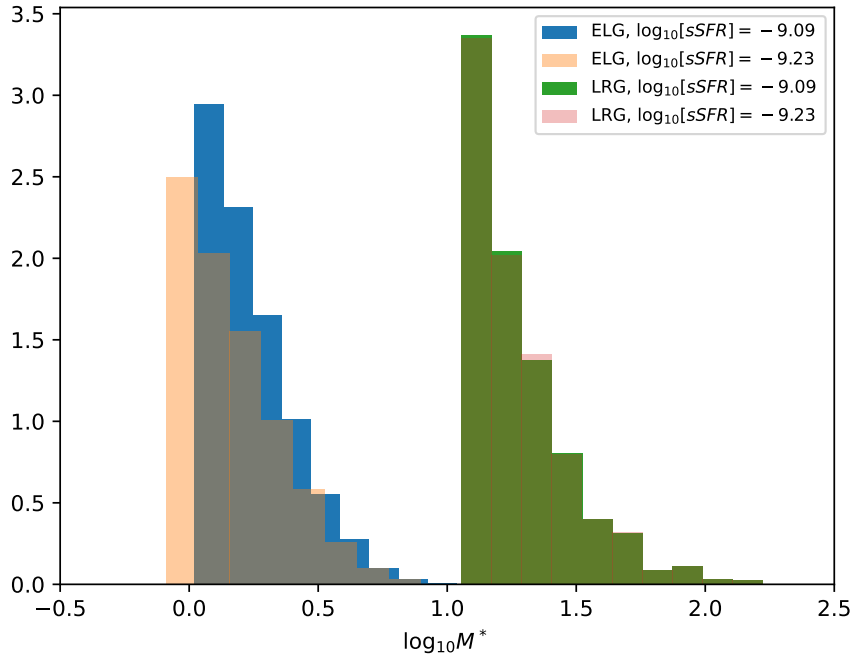
Figure 4.3: The stellar mass of ELG and LRG-like galaxies selected by two different sSFR cuts. LRGs are significantly heavier in both cuts, and it is apparent that the ELG and LRG selected samples do not overlap. While the different sSFR cuts significantly affect the M^* of the ELGs, they have almost no influence on the M^* of LRGs. The histograms are normalized for easier visualization, as ELG and LRG samples do not contain the same number of galaxies.

and used several different ML algorithms in combination with unstructured input features. The main reason for this is that GNN and CNN would generally need a larger data sample size, with higher number densities. We further justify this selection when we discuss the interpretability of our model in Section 4.3.

Input features

Each galaxy belongs to exactly one halo whose b_ϕ we want to infer. However, it is not necessarily true that the halo's center coincides with this galaxy's position. We define a central galaxy as the galaxy with the largest M^* within $R = 1\ h^{-1}$ Mpc of the galaxy whose halo properties we want to infer. R is a hyperparameter of our model, which we optimize as discussed in Section 4.3. We then use the central galaxy surroundings to describe the halo environment more accurately.

We use 13 features that we found useful for predicting b_ϕ:

- r, g, z – AB magnitudes of r, g, and z bands of the central galaxy

- $\log M^*$ of the central galaxy

- N_R; number of neighbor galaxies within R, for $R \in \{0.5, 1, 2, 3, 4, 5\}$ h^{-1} Mpc

- Metric of anisotropy, defined as $A = \left\| \sum_i \frac{\mathbf{R_i}}{\|\mathbf{R_i}\|} \cdot H(R_H - R_i) \right\|_2$

- Sum of stellar masses of the galaxies in the neighborhood; $\log\left(\sum_i M_i^* \cdot H(R_H - R_i)\right)$

- Sum of stellar mass to distances ratios; $\log\left(\sum_i \frac{M_i^*}{R_i} \cdot H(R_H - R_i)\right)$.

Here $\mathbf{R_i}$ is the separation vector between the i-th neighbor galaxy and central galaxy in redshift space, and $H(R_H - R_i)$ is the Heaviside step function that ensures that only galaxies within the radius R_H are included in the sum. R_H is a hyperparameter of our model, for which we use $R_H = 5\ h^{-1}$ Mpc. The results are mostly insensitive to the choice of R_h, however when considering a neighborhood much more distant than $5\ h^{-1}$ Mpc, the results do not improve, and rather start to deteriorate. The machine learning models can in theory learn to ignore less important input features, however, we are limited by the size of our dataset. For this reason, using fewer, only dominantly important features turns out to be better. All the described features use the distances calculated in redshift space. The predictive power of the model would increase if the features are calculated in real space, however we do not focus on them, since spectroscopic surveys observe galaxies in redshift space.

The correlations between the most representative input features are presented in Fig. 4.4. The intrinsic features of the galaxies are strongly correlated, with the correlation between luminosity bands being almost exactly 1. For this reason, we tried performing principal component analysis (PCA), which would reduce the dimensionality of the dataset. However, the results did not improve, and we do not use PCA to produce final predictions.

The stellar mass of the central galaxy, sum of neighborhood stellar masses, and sum of stellar-mass-to-distance ratios are all logarithmized before being used for machine learning. They are approximately exponentially distributed, therefore very large (small) numerical values could lead to potential numerical instabilities during the ML learning process. Furthermore, we standardize all features, $\hat{x}_i = \frac{x_i - \mu_i}{\sigma_i}$. Normalizing instead of standardizing features lead to slightly worse results, probably due to the outliers present.

We experimented with utilizing alternative input features. We already commented on the rationale behind restricting the input information to neighboring galaxies within a radius of R_H. However, it would be possible to instead select the N nearest galaxies with the objective of optimizing N. This approach appears to be less intuitive from a physical standpoint, as it may not accurately capture the influence of clustering effects. We have indeed checked and observed that it leads to slightly worse predictions for b_ϕ. Furthermore, we have checked and observed that incorporating information on the luminosities and stellar masses of the neighboring galaxies does not lead to any discernible improvements in the obtained predictions.

In Table 4.1, we have 8 different datasets for which we want to make separate b_ϕ predictions. When we make predictions for ELG galaxies, we also include information about
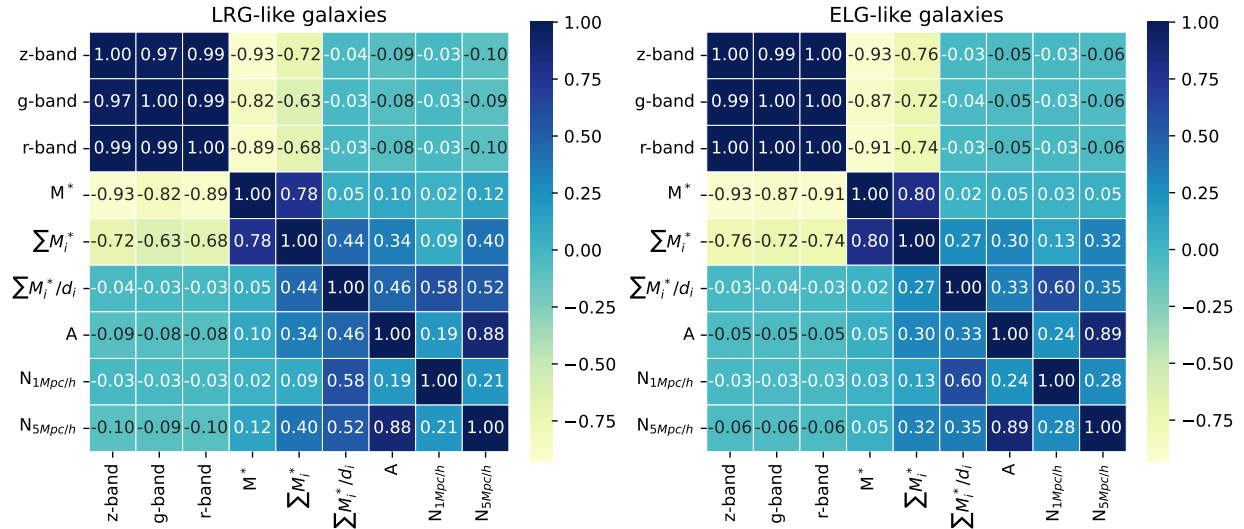
Figure 4.4: Pearson correlation coefficients of the most representative input features. The coefficients are rounded and are not exactly 1 between different luminosity bands. While the intrinsic features of galaxies (luminosity bands and stellar mass) are quite strongly correlated, they are mostly uncorrelated with the distributions of the galaxies in the neighborhood.

nearby LRG galaxies (in redshift-space) and vice versa. In other words, we use both ELGs and LRGs to calculate the number of neighbor galaxies or local anisotropy. However, we train the model and make predictions only on one type of galaxy at a time - this procedure could be performed on real data.

Models

We have tried using several different machine learning algorithms to predict b_ϕ from the input features as described in Section 4.3. The top three best-performing models are Artificial Neural Networks (ANN), Support Vector Machines (SVMs), and gradient-boosted tree models (XGBoost [329]). We discuss the process of hyperparameter optimization in Section 4.3.

We employ a technique known as model stacking to enhance the predictive performance of the three best-performing models; ANN, SVM, and XGBoost. Model stacking integrates individual predictions of multiple models and produces the final output from them. By stacking models, we can reduce both prediction bias and variance. The prediction errors arising from the individual models stem from distinct sources and thus can be partially offset.

The effectiveness of model stacking relies heavily on the diversity and quality of the base learners (the models whose predictions we are combining). A decrease in variance and new sources of prediction bias may be introduced if the base models are not good enough. ANN,

XGBoost, and SVM all perform well independently and have distinct modeling techniques and assumptions (XGBoost is an example of a tree-based ensemble method, ANN is a type of neural network, and SVM is a discriminative model). We tried including additional models in the ensemble, which did not lead to an increase in predictive power.

After obtaining the predictions of the base models, we want to train a new meta-model on them and map "base predictions" to b_ϕ. We want to find a function $f : \mathbb{R}^3 \to \mathbb{R}$, which takes three continuous inputs (predictions of base learners) and outputs b_ϕ. We use a simple ridge regression for this task to avoid overfitting.

Training

To optimize the hyperparameters of the algorithms described in Section 4.3, and features described in Section 4.3, we perform a train-validation-test split of the dataset (of each galaxy sample). The test set is held aside and is only used for the final evaluation. We use 3-fold cross-validation for training the meta-model.

Regarding the size of the train-test-validation split, we use a ratio of 70-20-10 %. Because we are dealing with spatially distributed galaxies, having a random split would lead to data leakage. The galaxies in the test (or validation) set may lie nearby the galaxies from the training set and would thus be in an identical environment. It would allow the model to perform well on the test set by simply memorizing the patterns from the training set, but would not be able to generalize well. We adopt the solution of splitting the data into spatially distinct regions based on x and y spatial coordinates, which are not the line-of-sight directions.

Using the training and validation sets, we employ grid search to tune the hyperparameters. We also evaluate the performance of other combinations of input features, different options of feature transformations, different loss functions, different models, and the performance of adopting a multi-target regression problem and predicting halo mass and concentration over b_ϕ.

Examples of features that we optimized have already been mentioned in the section 4.3. The optimizations include adjusting the distances to capture information about the surroundings and modifying the number of features that describe environmental density.

While optimizing the models, we explored a range of parameters. However, the optimizations improved the results only to some extent because the correlation between the input features we use and the halo assembly bias is limited (in addition to being complex).

Regarding the artificial neural networks (ANN), we have tried a range of hidden layers between 2 and 10, the number of neutrons per layer between 5 and 50, and we also changed the dropout rate and the activation functions. Our final ANN architecture has four hidden layers with 20 neurons per layer. In addition, we use dropout and the LeakyReLU activation function. The kernel of the SVM model used is the radial basis function. All other values of the hyperparameters and the values in the grid search grid are available on GitHub[2].

---

[2]`https://github.com/jmsull/ml_fnl_forecast`

Results
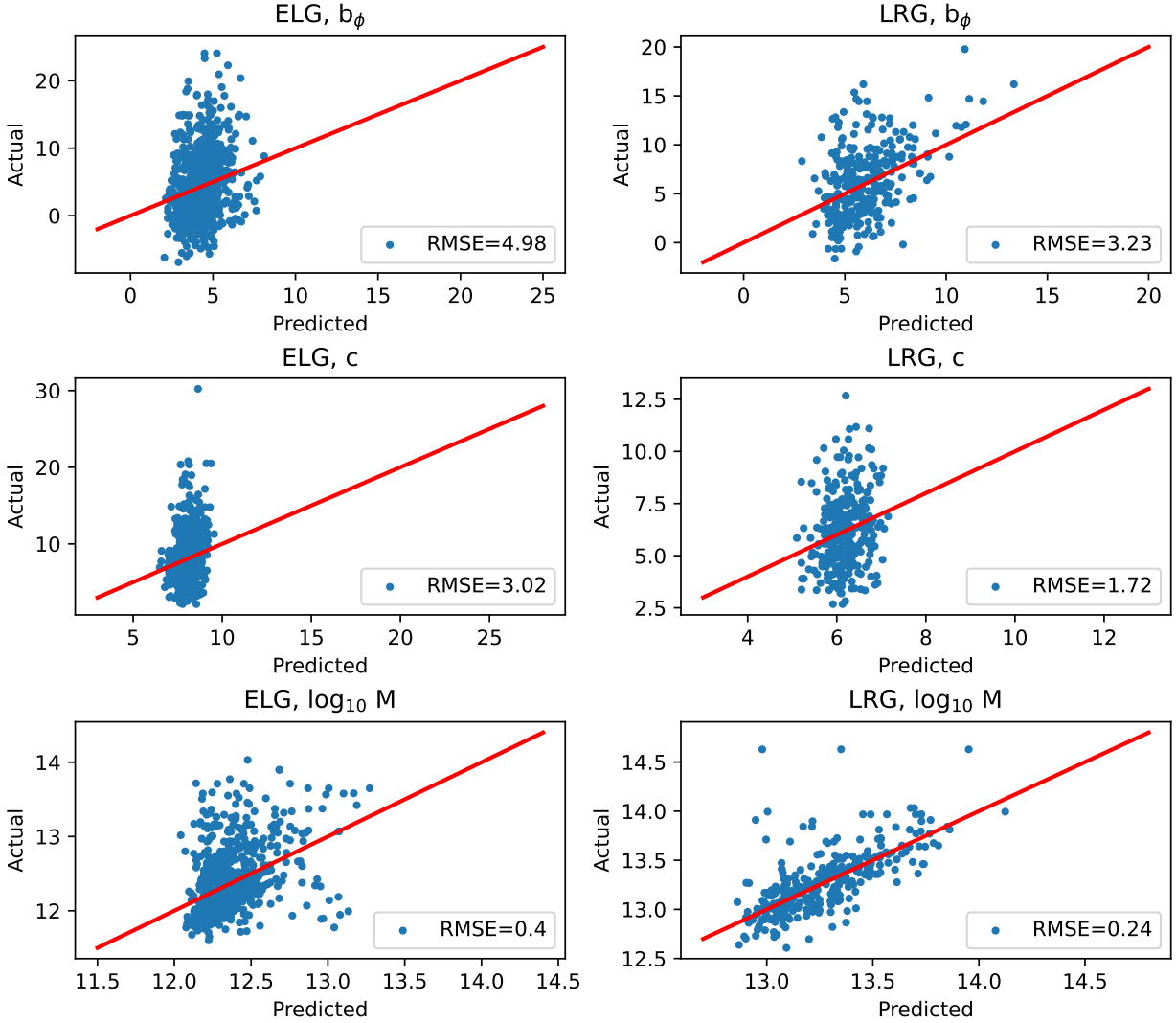
Figure 4.5: Results for ELG (left) and LRG (right) predictions on the test set. We present predicted versus actual values for each galaxy. The perfect predictions should lie on the red line, which has a slope of 1. Only the b_ϕ predictions (top row) are relevant for the analysis. However, we also show halo concentration and mass predictions (bottom two rows).

The final predictions made by the stacked model on the test set are presented in Fig. 4.5.. We also show the separate predictions for halo concentration and mass. We do not use mass and concentration predictions elsewhere, but it is useful to see them since b_ϕ is a function of these variables. We can see that for both ELGs and LRGs, halo mass predictions are significantly better than both b_ϕ and concentration predictions. While the

b_ϕ and concentration RSME are of the same magnitude, the RMSE of mass prediction is roughly 10 times smaller.

It is hard to make direct comparison between ELG and LRG results, since the distributions of the target values are different, as seen in Fig. 4.2. However, it seems that the results in terms of RMSE are generally better for LRGs. The reason for this could be that, as seen in Fig. 4.2, LRGs live in more massive halos, which are correlated with larger environments and higher central fractions. Moreover, LRGs are older than ELGs and are thus less affected by hydrodynamical effects, which are hard to capture with the properties we are looking at.

In Fig. 4.6 we can see the average b_ϕ values of the tertiles, split based on predicted b_ϕ from Fig. 4.5. The "Ideal" dots in Fig. 4.6 correspond to the values of b_ϕ for splits made by actual b_ϕ values of the galaxy host halos. After these tertilles are defined we compute the mean b_1 of all halos in each tertile.
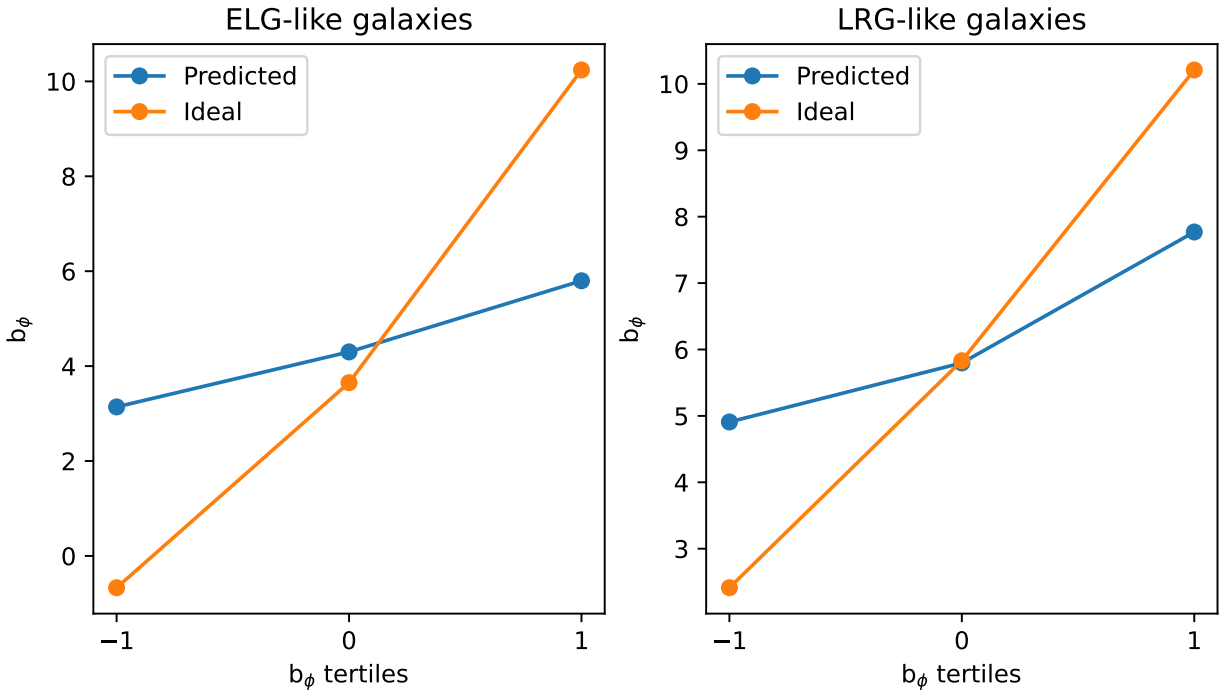
Figure 4.6: $\bar{b}_\phi$, for ELG and LRG samples. We split the sample into tertiles for the "Ideal" case based on the actual b_ϕ value and report the average b_ϕ for each tertile. For the "Predicted" case, we make the split based on predicted b_ϕ, as shown in Fig. 4.5. The horizontal axis on the plots is arbitrary and just denotes different tertiles.

We want to look at the interpretability of our model and see which information is most important for the obtained results. For this purpose, we use Shapley Additive explanations (SHAP) values, a technique for determining feature importance [330]. The SHAP feature importance is calculated by permutating all features and summing the average contribution

(a) ELG-like galaxies   (b) LRG-like galaxies
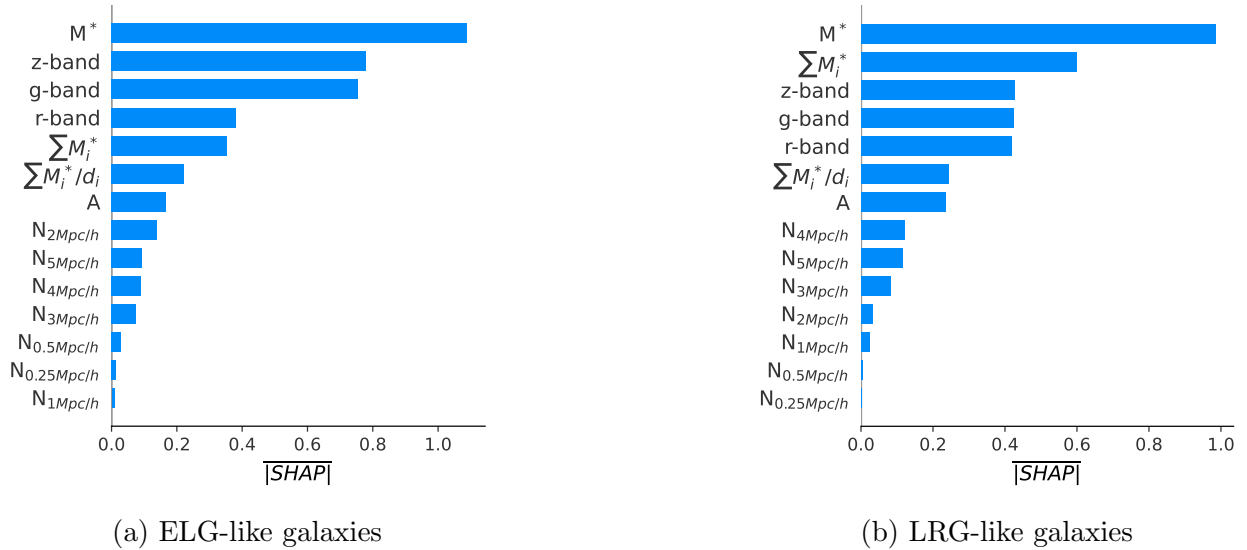
Figure 4.7: SHAP values of feature importance for predictions in redshift space. We can see that for both ELG and LRG- like galaxies, intrinsic features of galaxies are of much greater importance than the properties of the neighborhood. We can also observe that the feature importance is similar for both ELGs and LRGs.

of each feature over all permutations. It provides a fair and accurate estimation of the contribution of each feature, even if features are correlated, and can account for complex interactions between features[3].

The average SHAP values for each feature are presented in Fig. 4.7. The feature importance is presented for only two galaxy samples, however it remains practically the same for others. We can notice that, for both ELG and LRG, the intrinsic properties of the galaxies are significantly more important than environmental properties. Another interesting thing to observe is that the feature importance is similar for ELG and LRG galaxies.

We can observe that the importance of N_R does not converge to 0. This makes sense since there is a high correlation between the number of neighbors within different R. Moreover, we have explicitly checked that adding information about a larger environment does not improve results. It just takes some importance from other N_R features, while their sum would remain the same. The reasons are that the information too far from the halo does not affect the results, the dataset size is inadequate, and the correlation between halo properties and those features is noisy.

In Fig. 4.7, we can see that the information from the local neighborhood is much less influential than other features are when making predictions. Furthermore, we observed that adding information about luminosities and stellar masses of nearby galaxies does not improve

---

[3]The more popular technique used is "permutation importance", which has the main drawback of being unable to deal with correlated features.

the results. We can conclude that it thus makes sense to approach the problem with the unstructured data instead of using GNN or CNN. Those could work much better in real space, with a larger simulation (larger data size) and higher number densities.

4.4 Multi-tracer $f_{NL}^{\rm loc}$ forecasts

In this Section we provide Fisher forecast results for $\sigma(f_{NL}^{\rm loc})$ using a slightly extended version of the multi-tracer approach of Ref. [275]. In particular, we first provide forecasts for DESI LRGs and ELGs using the $b_\phi(b_1)$ relations found in Section 4.3 for mock versions of these samples. We then provide more speculative forecasts for future planned spectroscopic surveys (MegaMapper [331], SPHEREx [332]) that will target LPNG.

Forecast setup

Following [275], we perform forecasts of idealized constraints on $f_{NL}^{\rm loc}$, slightly generalizing the expressions there to a 3-tracer analysis. To be consistent with IllustrisTNG, we will use the Planck15 fiducial flat ΛCDM cosmology with $\Omega_m = 0.3089$, $\Omega_b = 0.0486$, $\Omega_\Lambda = 0.6911$, $h = 0.6774$, $\sigma_8 = 0.8159$, and $n_s = 0.9667$ in these forecasts [333, 334], and do not marginalize over ΛCDM parameters. While this choice might be considered optimistic, the multi-tracer method is relatively insensitive to marginalizing over ΛCDM parameters (see Appendix C of Ref. [335]). Furthermore, even without the multi-tracer method, we find that our single-tracer forecasts produce $\sigma(f_{NL}^{\rm loc})$ that is 10% lower than the (marginalized) values quoted in Ref. [269] or less (see Fig. 14 there) except for DESI ELGs, where the difference is 25%, though in this case, we do not use the same redshift distribution or redshift range as in Ref. [269]. We will work entirely in linear theory at the power spectrum level (including linear redshift space distortions $b \to b + f\mu^2$), and use a simple linear bias model only. We do not account for the Alcock-Paczynski effect [336]. We compute the linear matter power spectrum using `camb` [337]. We use 101 linear μ bins between 0 and 1, $N_z = 65$ linear z bins between $z_{\rm min}$ and $z_{\rm max}$, and 100 k values between $k_{\rm min}$ (which is survey dependent) and 0.075 $h/$Mpc. We verified that doubling the number of z or μ bins or k points beyond these values changes the final forecasted error for the single-tracer forecasts by less than 1%. To be consistent with Ref. [269], we perform all single-tracer forecasts with the UMF prediction with $p = 1$. For a more detailed discussion of a careful treatment of the choice of p in the context of real data, including the choice of priors, we refer the reader to Ref. [290]. We neglect wide-angle effects and large-scale relativistic effects on the galaxy power spectrum. We neglect any additional systematic effects that could contaminate a large-scale measurement of the power spectrum (e.g. stellar contamination), though it is of course extremely important to characterize these well in a real data analysis.

Here we describe the galaxy samples used in our forecasts. The primary samples we consider are the Dark Energy Spectroscopic Instrument (DESI) Luminous Red Galaxies (LRGs) and Emission Line Galaxies (ELGs). For the DESI LRG-like sample we use the redshift dis-

tribution of [338], and for the DESI ELG-like sample we use the redshift distribution of [320]. For DESI $f_{\mathrm{sky}} = 0.34$ [312], $k_{\mathrm{min}} = 0.0023\ h/\mathrm{Mpc}$, and $z_{\mathrm{min}} = 0.0$, $z_{\mathrm{max}} = 2.0$. For the linear galaxy bias of the two DESI samples in the single-tracer analysis, we make the empirically calibrated choice [312, 339] of $b_{\mathrm{LRG}} = \frac{1.7}{D(z)}$, $b_{\mathrm{ELG}} = \frac{0.84}{D(z)}$, where $D(z)$ is the linear growth factor (normalized to 1 today) computed in our fiducial cosmology. For the multi-tracer forecasts using the results of Section. 4.3, we assume the results evolve similarly with $D(z)$ (i.e. $b(z) = \frac{D(z=1)}{D(z)}b$, and similarly for b_ϕ).

For the MegaMapper forecasts, we follow the prescriptions of Refs. [340, 269, 331, 341] for the galaxy density $n(z)$ and linear bias $b(z)$, we assume $f_{\mathrm{sky}} = 0.34$, $k_{\mathrm{min}} = 0.0017\ h/\mathrm{Mpc}$, and $z_{\mathrm{min}} = 2.0$, $z_{\mathrm{max}} = 5.0$. For SPHEREx, we use $f_{\mathrm{sky}} = 0.65$, $k_{\mathrm{min}} = 0.001\ h/\mathrm{Mpc}$, and $z_{\mathrm{min}} = 0.1$, $z_{\mathrm{max}} = 3.0$. For the SPHEREx forecasts, we assign linear bias as a function of redshift based on the procedure outlined in Ref. [332], but use the fitting functions of Ref. [342] as implemented in COLOSSUS [343] (and have verified that this choice changes the halo bias of individual redshift error samples by less than 1% with respect to the fitting functions of Refs. [344, 345]). We consider the redshift-error samples (labeled by $\frac{\sigma_z}{1+z}$) on their own and in combination. We incorporate the SPHEREx redshift errors as described in Ref. [332] for the various redshift error samples, multiplying the shot noise term by $\exp(-\left[k\mu\sigma_z\frac{d\chi}{dz}\right]^2)$. All non-SPHEREx forecasts in this paper assume zero redshift error. When combining several redshift error samples, we always use the largest redshift error. For the fully-combined SPHEREx sample, this corresponds to an assumed redshift error of $\frac{\sigma_z}{1+z} = 0.2$, so we deviate slightly from the treatment of Ref. [269], who used $\frac{\sigma_z}{1+z} = 0.05$ for this full sample, and therefore we are more pessimistic in our forecasts. We compute an effective bias for combined samples by weighting the bias by the sample number density.

We work in linear theory and at a fiducial cosmology where $f_{NL}^{\mathrm{loc}} = 0$, and so use the Fisher matrix for $\delta_X(k, \mu, z)$ which, in the Gaussian approximation is [346]:

$$F_{\alpha\beta} = \sum_{i_z=1}^{N_z-1} \int_0^1 d\mu \int_{k_{\mathrm{min}}}^{k_{\mathrm{max}}} N_{k,i_z} \mathrm{Tr}\left[\mathbf{C}_{,\alpha}\mathbf{C}^{-1}\mathbf{C}_{,\beta}\mathbf{C}^{-1}\right] \tag{4.6}$$

where $\mathbf{C}$ can be either the two-tracer version $\mathbf{C}^{(2)}$ or the three-tracer version $\mathbf{C}^{(3)}$, with

$$\mathbf{C}^{(2)} = \begin{pmatrix} C_{XX} & C_{XY} \\ & C_{YY} \end{pmatrix}, \quad \mathbf{C}^{(3)} = \begin{pmatrix} C_{XX} & C_{XY} & C_{XZ} \\ & C_{YY} & C_{YZ} \\ & & C_{ZZ} \end{pmatrix}, \tag{4.7}$$

where the field covariance and number of modes N_{k,i_z} in a volume shell V_{i_z} are

$$C_{XY} = \left(b^X + \mu^2 f + b_\phi^X f_{NL}^{\mathrm{loc}}\mathcal{M}^{-1}(k)\right)\left(b^Y + \mu^2 f + b_\phi^Y f_{NL}^{\mathrm{loc}}\mathcal{M}^{-1}(k)\right)P_L(k) + \frac{\delta_{XY}^D}{\bar{n}_X}, \tag{4.8}$$

$$N_{k,i_z} = \frac{dk k^2 V_{i_z}}{2\pi^2}, \tag{4.9}$$

respectively, and we neglect both tracer non-Poisson stochasticity and cross-stochasticity between different tracer fields [347, 335, 348]. Here i_z gives the index of the $z-$bin midpoint used to compute Fisher matrix element, and V_{i_z} is a spherical shell computed using the adjacent redshift bin edges and is scaled by f_{sky}.

Here, in the notation of Ref. [275], $\alpha = \frac{b_1}{b_2}$ and $P_2 = b_2^2 P$ (at our fiducial value of $f_{NL}^{\text{loc}} = 0$), where b_1 and b_2 are the *linear* biases of the first and second samples under consideration. We define $\beta \equiv \frac{b_3}{b_2}$ analogously for the three-tracer case. Explicit expressions for the Fisher matrix element $F_{f_{NL}^{\text{loc}} f_{NL}^{\text{loc}}}$ are provided in Appendix 4.6 and in the public code accompanying this paper (linked above).

Results for $\sigma(f_{NL}^{\text{loc}})$

DESI

Sample	$\sigma_{\text{ST}}(f_{NL}^{\text{loc}})$	$\sigma_{2,\text{MT}}(f_{NL}^{\text{loc}})$	$\sigma_{2,\text{MT}}^{(P)}(f_{NL}^{\text{loc}})$	$\sigma_{3,\text{MT}}(f_{NL}^{\text{loc}})$	$\sigma_{3,\text{MT}}^{(P)}(f_{NL}^{\text{loc}})$
ELG, ideal	7.1	1.3	1.4	1.7	1.5
ELG, pred.	.	10	2.5	12	2.3
LRG, ideal	5.0	3.0	2.6	3.5	3.4
LRG, pred.	.	16	3.7	17	4.7

Table 4.2: Local primordial non-Gaussianity forecasts for single- and multi-tracer forecasts for ELG and LRG-like simulated galaxies. For each column $\sigma_{i,XT}^{(P)}$, i denotes the number of tracers used in each forecast, (P) indicates whether the forecast includes P_2 as a parameter, and $X = \{S, M\}$ is S for single-tracer and M for multi-tracer. For the 2-tracer forecasts, we use the highest and lowest b_ϕ tertiles.

Here we present $\sigma(f_{NL}^{\text{loc}})$ forecast results - we first perform forecasts for LRGs and ELGs individually with several multi-tracer setup choices (Table 4.2), then present results illustrating the effect of selection choice on $\sigma(f_{NL}^{\text{loc}})$ for ELGs (Table 4.3) before finally considering the case of ELGs and LRGs in their overlap region (Table 4.4).

Table 4.2 shows the improvement of the multi-tracer forecasts over the single-tracer forecasts for several choices of multi-tracer setup for our fiducial choice of DESI mock galaxy selection ($\bar{n}_{\text{ELG}} = 5 \times 10^{-4}$ $[h^{-1}\ \text{Mpc}]^{-3}, \log_{10}\left(\frac{sSFR}{h\ yr^{-1}}\right) = -9.09$). Here we report results for both the simulated LRG and ELG galaxy samples, for the ideal values of $b_\phi(b)$ in each tertile (i.e. concentration information is recovered perfectly) and for the predicted ("pred.") values using the machine learned relationship $b_\phi(b)$ from observable galaxy properties. While clearly the ML model cannot access all the f_{NL}^{loc} information contained in halo mass and concentration, it recovers enough information to significantly improve the forecasted LPNG amplitude error $\sigma(f_{NL}^{\text{loc}})$. The improvement for the predicted relations is largest for ELGs,

where we can obtain a forecasted error of $\sigma(f_{NL}^{\text{loc}}) = 2.3$, a factor of 3 improvement over the single-tracer case.

We show both 2-tracer and 3-tracer forecasts. The 2-tracer forecasts use the upper and lower tertiles of the three-bin splits, while the 3-tracer forecasts use all three tertiles (both cases use number density $\bar{n}/3$ for each tracer) When making multi-tracer forecasts, we compute the error on f_{NL}^{loc} using two different procedures - first, when the only parameters considered in the Fisher matrix is (are) the relative amplitude(s) α, (β), and second, when both the relative amplitude(s) α, $(\beta,)$ and P_2, the power spectrum corresponding to the second sample are used. Since in the first case we use only the relative amplitude for the two-tracer forecasts, we refer to the case of including P_2 as a parameter as "(P)", since we are including the power spectrum as a parameter. Including P_2 as a parameter generally reduces the forecasted error, however, in the sample-variance limit P_2 adds no information (see discussion around eqn. 4.15). In multi-tracer forecasts used in the rest of this paper, we include P_2 as a parameter. Some brief further discussion of these forecasting aspects can be found in Appendix 4.6.

$\log_{10}\left(\frac{sSFR}{h\ \text{yr}^{-1}}\right)$	\-9.09		\-9.23	
$n_{\text{ELG}}\ [h^{-1}\ \text{Mpc}]^{-3}$	ideal	pred.	ideal	pred.
5×10^{-4}	1.5	2.3	1.4	2.3
7×10^{-4}	1.6	2.3	1.5	2.2
1×10^{-3}	1.6	2.5	1.5	2.2

Table 4.3: Multi-tracer $\sigma(f_{NL}^{\text{loc}})$ Fisher forecasts for several choices of $\bar{n}_g$ and sSFR for DESI ELG-like simulated TNG galaxies. Results do not vary significantly with the choice of mock galaxy selection.

Table 4.3 characterizes the effect of the simulated galaxy sample selection strategy on the forecasts for $\sigma(f_{NL}^{\text{loc}})$. These selections were mentioned already in Section 4.3 (in the context of the learned $b_\phi(b)$ relation), but here we supplement this by checking the effect of the choice of specific star formation rate split and stellar mass threshold, which determines the number density for a fixed sSFR split, on the final forecasted f_{NL}^{loc} error. We find that the final results are mostly insensitive to these choices, as they vary only by a factor of 17% or less (within the ideal or predicted cases). It may appear counter-intuitive that $\sigma(f_{NL}^{\text{loc}})$ increases even in the ideal case for both specific star-formation rates as the number density increases, but this is due to the fact that the mean linear bias and, at first approximation, the LPNG bias both shift downward as the number density increases. In more detail, while the b_ϕ values drop uniformly across all tertiles with the mean, the b value in the highest tertile hardly changes with increasing number density, but the lower tertile drops significantly, and this leads to a lower change in linear bias between the upper and lower tertiles at higher number density, which in turn leads to a smaller relative amplitude Jacobian factor in the Fisher matrix (see

eqn. 4.10). In any event, the highest difference of 17% in $\sigma(f_{NL}^{\rm loc})$ between the two choices of sSFR is for the highest number density sample (which is not used elsewhere).

ELG + LRG	$\sigma(f_{NL}^{\rm loc})$
$p = 1$	4.0
(2) ($\overline{\rm LRG}$, $\overline{\rm ELG}$), ideal	2.3
(2) ($\overline{\rm LRG}$, $\overline{\rm ELG}$), pred	2.3
(2) (LRG+, ELG+), ideal	1.4
(2) (LRG+, ELG+), pred	2.4
(3) (LRG-, LRG+, ELG-), ideal	0.8
(3) (LRG-, LRG+, ELG-), pred	2.0
(3) (LRG-, ELG+, ELG-), ideal	0.8
(3) (LRG-, ELG+, ELG-), pred	2.0
(3) (LRG-, ELG+, else), ideal	0.6
(3) (LRG-, ELG+, else), pred	1.5

Table 4.4: Local primordial non-Gaussianity amplitude $f_{NL}^{\rm loc}$ multi-tracer forecasts for ELG-LRG in their overlap region. The entries listed "($\overline{\rm LRG}$, $\overline{\rm ELG}$)" use the mean $b_\phi(b)$ prediction with the full number density of each sample Here "+" and "-" denote the top and bottom tertile values of $b_\phi(b)$ of the respective samples, respectively. The row entries following the ($\overline{\rm LRG}$, $\overline{\rm ELG}$) rows use (i)-tracer forecasts ($i = 2, 3$) with number densities of $\bar{n}/3$ for each sample. In the final two rows, the unused subsamples are combined with bias weighted by their number densities.

Table 4.4 shows $\sigma(f_{NL}^{\rm loc})$ Fisher forecasts for several multi-tracer sample configurations drawn from the ELG and LRG simulated galaxy samples. To represent the simplest possible multi-tracer forecast setup using the undivided LRG and ELG samples, we report $\sigma(f_{NL}^{\rm loc}) = 4.0$ for the UMF bias prediction (eqn. 4.2) with the parameter p set to 1. This forecast does not make use of the learned $b_\phi(b)$ relations presented in this paper. We next consider a similar case that uses the mean of the $b_\phi(b)$ values for the simulated LRG sample and the simulated ELG sample (with the fiducial choice of sample parameters: $\bar{n}_{ELG} = 5 \times 10^{-4}$, $\bar{n}_{LRG} = 2 \times 10^{-4}$, and $\log_{10}({\rm sSFR}) = -9.09$).

When we consider the multi-tracer forecasts that use two samples split by b_ϕ tertile, we find a lower $\sigma(f_{NL}^{\rm loc})$. When using the upper tertile samples for both ELGs and LRGs (LRG+,ELG+), we find a factor of 2.9 reduction in $\sigma(f_{NL}^{\rm loc})$ compared to the $p = 1$ UMF forecast in the ideal case, where concentration and halo mass information are perfectly recovered, and a factor of 1.7 reduction in $\sigma(f_{NL}^{\rm loc})$ when using the learned relation. We also consider three-tracer forecasts for the ELG and LRG tertile samples. Using the lower and upper tertiles for the LRGs and lower tertiles for the ELGs (LRG-,LRG+,ELG-), we find smaller forecasted errors than in the two-tracer case, and using the lower and upper tertiles for the ELGs along with the lower tertile for the LRGs (LRG-,ELG+,ELG-) provides similar

results, reducing the forecasted error by a factor of 5 compared to the $p = 1$ UMF forecast in the ideal case and a factor of 2 in the predicted case.

Finally, based on the promising results of the (LRG-,ELG+,ELG-) choice of subsamples, we show in the last two rows of Table 4.4 forecasts for the case where ELG- in this triplet is replaced with all the other sub-samples combined. In particular, the third sample is the combination of ELG-, ELG0, LRG0, and LRG+, where "0" indicates the central tertile. Bias values for these samples are weighted by their number density (as for the SPHEREx redshift error $\frac{\sigma_z}{1+z}$ samples). In this case, we find the most constraining forecasts we report for LRGs and ELGs, with $\sigma(f_{NL}^{\rm loc}) = 0.6$ and $\sigma(f_{NL}^{\rm loc}) = 1.5$ for the ideal and predicted b_ϕ values, which are now factors of 6.7 and 2.7 reductions over the naive $p = 1$ multi-tracer forecast, respectively.

We do not list all possible permutations of the ELG and LRG tertile samples (and their combinations) for brevity, though some of the other combinations produce a lower $\sigma(f_{NL}^{\rm loc})$ in the predicted case than we show here. It would also be interesting to further consider optimally combining split subsamples in general [e.g. as for mass in 349, 335] - we took a first step in exploring this here by using the "else" samples in Table 4.4, which produces excellent results.

We now discuss the origin of the sizeable reduction in our forecasts for $\sigma(f_{NL}^{\rm loc})$ over the single-tracer case. The Jacobian factor involved in converting from the Fisher information on the relative amplitude multi-tracer parameters (F_{ij} where $i, j \in \{\alpha, \beta\}$) to that of LPNG ($F_{f_{NL}^{\rm loc} f_{NL}^{\rm loc}}$) involves the factor (reproduced from Appendix 4.6):

$$\frac{\partial \alpha}{\partial f_{NL}^{\rm loc}} = \alpha \left(\frac{b_{\phi,1}}{b_1} - \frac{b_{\phi,2}}{b_2} \right) \mathcal{M}^{-1}(k), \tag{4.10}$$

which is evaluated at $f_{NL}^{\rm loc} = 0$. For a fiducial choice of $f_{NL}^{\rm loc} = 0$, the Jacobian factor of eqn. 4.10 is in fact the only place that the LPNG bias enters the Fisher forecast calculation. A similar expression holds for $b_3, b_{\phi,3}$ and β in the 3-tracer case.

If we consider the limit where the relative amplitude $\alpha \to 1$, then clearly this factor is proportional to Δb_ϕ. This limit is exact in a 2-way split of a single halo mass bin if linear bias does not evolve across the bin or with the splitting parameter.

We now make the even further simplification, for illustrative purposes, that each b_ϕ of the two samples in question is described by a UMF relation for some choice of the parameter p, and so $b_{\phi,1} = 2\delta_c(b_1 - p_1)$, $b_{\phi,2} = 2\delta_c(b_2 - p_2)$, and for $\alpha \to 1$ we have $\Delta b_\phi \to 2\delta_c \Delta p$, and $\frac{\partial \alpha}{\partial f_{NL}^{\rm loc}} \to -2\delta_c \Delta p \mathcal{M}^{-1}(k)$. From this demonstration, it would appear that we can drive our forecasted error on LPNG as low as we wish as we take $\Delta p \to \infty$. We further discuss the implications of this apparent conclusion in the next section.

Future LPNG surveys

To explore the potential of using samples with large differences in b_ϕ further, we apply our multi-tracer forecasting framework to two future surveys that will target LPNG, MegaMapper [331], and SPHEREx [332] using the setup described in Section 4.4. We use the UMF

relations as input for $b_\phi(b_1)$ in these forecasts, as we do not have the information necessary to construct detailed simulated galaxies and corresponding learned $b_\phi(b_1)$ for the future MegaMapper or SPHEREx galaxy populations. Since we do not have a solid basis for selecting Δp for these forecasts, we instead consider forecasts at a range of Δp. For the following multi-tracer forecasts, we will set $b_{\phi,i} = 2\delta_c(b_i - [p \pm \Delta p])$ where $p = 1$ unless otherwise indicated[4]. The value $p = 1$ is the best choice to make contact with the literature, which mostly uses the UMF with $p = 1$. In principle, if we were to choose a lower p, we should only reduce $\sigma(f_{NL}^{loc})$ (see Fig. 4.8) - this is closely related to the zero-bias effect discussed by Ref. [350].
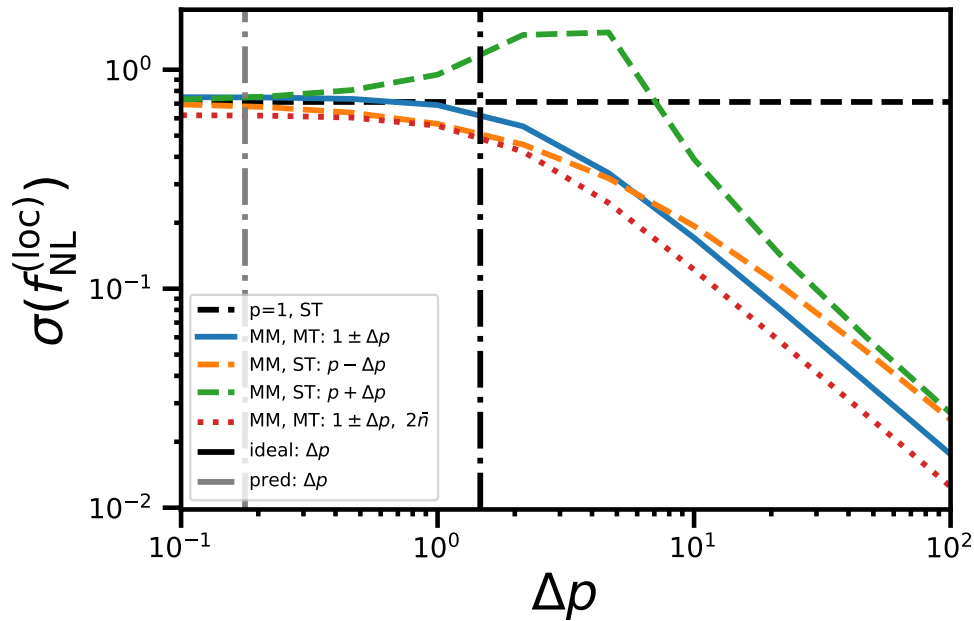
Figure 4.8: (*MegaMapper*): The dependence of forecasted $\sigma(f_{NL}^{loc})$ multi-tracer forecasts on Δp, the difference in the UMF parameter, for the planned MegaMapper survey. The black dashed line shows the single-tracer (ST) forecast with the original UMF scenario, for which $p = 1$. The solid blue line shows the multi-tracer (MT) forecast for the case in which the full MegaMapper sample is split into tertiles and the top and bottom tertiles are used with equal number density bins of $\frac{\bar{n}}{3}$, each of which is assigned $b_\phi = 1 \pm \Delta p$. For comparison, the dashed colored lines show the single tracer forecasts with $p = 1 + \Delta p$ (green dashed) and $p = 1 - \Delta p$ (orange dashed). The red dotted line is the same as the blue line (multi-tracer forecast), but with twice the number density. The horizontal dash-dotted lines in gray and black show the approximate values of Δp for the learned and ideal b_ϕ values for the mock DESI galaxy samples.

---

[4]This is in a sense a somewhat pessimistic choice, see Appendix 4.6 for more discussion
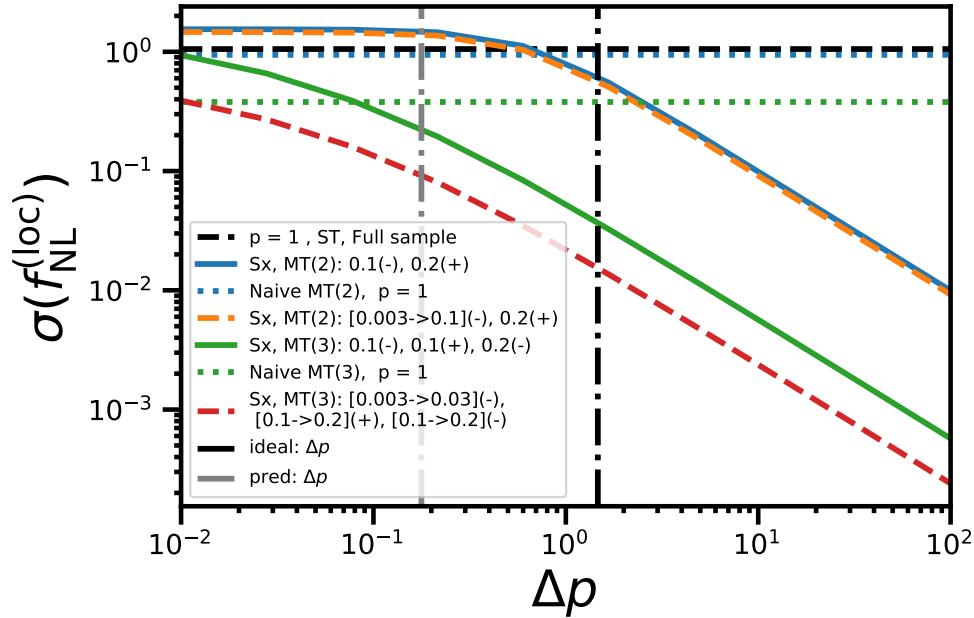
Figure 4.9: (*SPHEREx*): Similar to Figure 4.8, but for multi-tracer forecasts involving multiple SPHEREx redshift error samples. The black dashed line again shows the UMF single-tracer forecast. The solid blue line shows the two-tracer (2) multi-tracer (MT) forecast for the $\frac{\sigma_z}{1+z} = 0.01$ and $\frac{\sigma_z}{1+z} = 0.2$ redshift error samples, where the former has UMF parameter $p = 1 - \Delta p$ ("-") and the latter has $p = 1 + \Delta p$ ("+"). The orange dashed line shows the two-tracer MT forecast where the first tracer is the lowest redshift error sample ($\frac{\sigma_z}{1+z} = 0.003$) with $p = 1 - \Delta p$ and the second tracer is the combined sample of all other redshift error samples (with $p = 1 + \Delta p$). The green solid line shows the three-tracer (3) multi-tracer forecast for the three splits of the redshift error samples - the $\frac{\sigma_z}{1+z} = 0.01$ sample with $p = 1 - \Delta p$ (-), the $\frac{\sigma_z}{1+z} = 0.1$ sample with $p = 1 + \Delta p$ (+), and the $\frac{\sigma_z}{1+z} = 0.2$ sample with $p = 1 - \Delta p$ (-). The red dashed line shows another three-tracer (3) forecast using combined samples - the combined $\frac{\sigma_z}{1+z} = 0.003, 0.01$ samples with $p = 1 - \Delta p$ (-), the combined $\frac{\sigma_z}{1+z} = 0.03, 0.1, 0.2$ samples with $p = 1 + \Delta p$ (+), and the second half of the same sample with $p = 1 - \Delta p$ (-). Again, horizontal dash-dotted lines in gray and black show the approximate values of Δp for the learned b_ϕ for the mock DESI galaxy samples.

Figure 4.8 shows the resulting dependence of the two-tracer forecasted error $\sigma(f_{NL}^{\text{loc}})$ on Δp for MegaMapper. The solid blue line shows the two-tracer forecast for the scenario in which the number density of the MegaMapper sample is split into thirds, and the top and bottom tertiles of the sample are assigned the *same* linear bias b and an LPNG bias of either $b_\phi = 2\delta_c(b - (1 \pm \Delta p))$. When $\Delta p \to 0$, we find that we approach the single-tracer results at $p = 1$, (since for $\Delta p = 0$ the relative amplitude Jacobian factors vanish and we are left with only P_2 information). It is immediately apparent that there is no bound on how much the

multi-tracer forecast can improve upon the single-tracer forecast at fixed p.

To make a fairer comparison to the single tracer case, we also plot in the dashed lines the single tracer forecasts with $p = 1 + \Delta p$ (green) and $p = 1 - \Delta p$ (orange). We can see that by drastically increasing or decreasing p in the single-tracer case we can obtain a large reduction in $\sigma(f_{NL}^{\text{loc}})$ that is qualitatively similar to that of the multi-tracer case. In the limit of $\Delta p \to \infty$ we see that the multi-tracer case always outperforms the single-tracer case by a roughly constant factor, and both lines have similar slope. This is expected due to SV cancellation and the fact that (for fixed b) in the large-p limit, we have that $\sigma_{ST}(f_{NL}^{\text{loc}}) \propto \frac{b}{p}$, and for large Δp, $\sigma_{MT}^{(2)}(f_{NL}^{\text{loc}}) \propto \frac{b}{\Delta p}$. These single-tracer forecasts use the full number density of the sample, so to account for this difference when comparing to multi-tracer, we also artificially increase the number density by a factor of two in the multi-tracer case as an additional point of comparison (red dotted line). Here we can see that the slight improvement of single-tracer over multi-tracer around $\Delta p = 1$ is due to the reduced number density of the multi-tracer sample.

We also mark in vertical dash-dotted gray and dash-dotted black the values of Δp that are close to describing b_ϕ for the upper and lower tertiles of the ELG and LRG simulated galaxies in the predicted and ideal cases, respectively[5]. While this exercise illustrates in part where the improved constraints for DESI galaxies is coming from, the change in Δp is of course not the only thing driving a difference in the mulit-tracer and single-tracer forecasts. As in the usual halo-mass-based multi-tracer application, b_1 still plays a significant role in both the Jacobian factor of eqn. 4.10 and the Fisher matrix elements of the multi-tracer parameters (e.g. through factors of the form $b^2 P \bar{n}$). For anything but the idealized sample-variance limit, the increased number density from splitting the full sample can also strongly affect the final forecasted error. Further study of physical and practical limits on Δp will be the subject of future work.

We performed a similar exercise for SPHEREx in Figure 4.9, which shows forecasted $\sigma(f_{NL}^{\text{loc}})$ as a function of Δp for several combinations of SPHEREx redshift error samples. In black dashed we again show the single-tracer forecast for $p = 1$ for the combined sample of all SPHEREx redshift error samples (using the most pessimistic redshift error). Multi-tracer forecasts for individual redshift error samples (labeled by $\frac{\sigma_z}{1+z} \in \{0.003, 0.01, 0.03, 0.1, 0.2\}$) where we assign splits on Δp are shown by the solid lines. We also show "naive" two-tracer forecasts when considering different redshift error samples using the same p (but these samples have different b). To take advantage of the high number density of SPHEREx, we also show multi-tracer forecasts for combinations of redshift error samples (for example, combining the two lowest redshift error samples to create one multi-tracer sample and combining the three highest redshift error samples to create the second multi-tracer sample) as the dashed lines. Here the "MT(i)" and ($\pm$) labels serve the same purpose as in Table 4.4, indicating how many tracers are used in the forecast and whether the top or bottom tertile of the Δp-split sample is used.

---

[5]These values are simply calculated as $\Delta p = \frac{1}{2}(\delta^c)^{-1}(b_\phi^+ - b_\phi^-)$, so there is a symmetry here that isn't present in the ELG/LRG results. For more discussion see Appendix 4.6

We find similar results to Fig. 4.8, in that the multi-tracer forecasts improve over the single-tracer forecasts as Δp increases. First, we discuss the case where we use individual redshift error samples as the multi-tracer samples (solid lines). The naive forecasts (dotted lines) are also for the individual redshift samples, and should be compared to lines of the same color.

At low Δp the improvement of the two-tracer Δp multi-tracer forecast over the full single-tracer forecast (black-dashed) with the full sample disappears since both are dominated by the most pessimistic redshift error of $\frac{\sigma_z}{1+z} = 0.2$. This is in fact worse than the naive multi-tracer case (dotted blue line) due to the lower number density. Similar statements hold for the three-tracer case, though the three-tracer results show what is effectively only the high Δp regime of the 2-tracer case, and, we can see that the difference between the naive and Δp-split three-tracer forecasts is larger than in the two-tracer case.

Turning now to the combined samples, we see that combining several redshift samples (with the appropriated weighted linear biases) leads to a reduction in shot noise[6] due to increased number density of the sample, and this gives lower $\sigma(f_{NL}^{\mathrm{loc}})$ when $\Delta p \to 0$ as seen in the difference between solid and dashed multi-tracer lines. We can also see from the red dashed line that the improvement over single-tracer is similar to the red dotted curve in Fig. 4.8, again suggesting that the uniform reduction in $\sigma(f_{NL}^{\mathrm{loc}})$ over the three-tracer single sample forecast (green solid line) is due to the increased number density of the combined samples. Essentially all of the lowest error multi-tracer forecasts come from the high redshift error samples (or combinations that use the highest redshift error), as these have the highest number density (roughly $2 - 3 \times 10^{-3}$ $[h^{-1}\ \mathrm{Mpc}]^{-3}$), largest redshift range (extend out to $z = 3$), and the measurement of f_{NL}^{loc} is relatively insensitive to these errors.

Discussion

We are leveraging information about b_ϕ through the portal of halo mass and concentration, but in galaxy survey data we only have access to observable galaxy properties, not the concentration itself. We have fixed the galaxy-halo connection (GHC) that maps the halo field to the galaxy field for a specific galaxy sample to that afforded by the IllustrisTNG galaxy formation model. Even assuming this fixed GHC, we will inevitably lose information about the halo mass and concentration when passing to simulated galaxies. For this reason, we report multi-tracer forecasts for both "ideal" and "predicted" b_ϕ for the ELG and LRG galaxy samples, where the predicted case captures the limitations of using only observable galaxy properties rather than host halo properties. The former refers to the (fictitious) scenario in which we could perfectly recover b_ϕ from the simulated galaxy sample based on mass and concentration, and the latter to what the ML model of Section 4.3 recovers. The "ideal" forecasts results should therefore be interpreted as quantifying the maximal b_ϕ information available from the given galaxy sample, while the "predicted" forecasts tell

---

[6]The redshift error is unaffected, as we always choose the most pessimistic redshift error for a combined sample.

us what information about b_ϕ can be obtained by applying the ML predictions directly to DESI data (assuming the TNG galaxy formation model, and that the observable input features have actually been measured). Following this initial work, several strategies could be employed to improve the ML prediction of b_ϕ. The larger MilleniumTNG simulation [351] could be used, which would increase the size of the datasets used by a factor of almost 15. We also found, as expected, that results were much improved in real space compared to redshift space, so adding a reconstruction procedure [352] could potentially improve results. Another option would be to use additional galaxy properties that are more challenging to model as input features, such as stellar half-mass radius, metallicity, or maximum value of spherically-averaged rotation curve.

We have made several simplifying assumptions in this work when applying the $b_\phi(b_1)$ results of Section 4.3 to the Fisher forecasts of this section. We have ignored the impact of fiber collisions, which, for real galaxies, would impact the ability to determine close pairs of galaxies and to compute the local environment statistics used for determining b_ϕ from observables. However, we expect that since local environment features are significantly less informative than the magnitude bands and stellar mass, including fiber collisions would have a minimal impact on our results. We used the concentration-LPNG bias relation of Ref. [296] which uses different simulations than those of IllustrisTNG itself, but since Ref. [296] uses logarithmized concentration, we do not expect a significant effect from this choice. We neglected the exact redshift dependence of the linear bias b, instead using the linear growth factor to account for the redshift dependence as in Ref. [312]. While here we only worked with halo concentration and mass, there is certainly no a priori reason to ignore halo properties such as age (e.g. vis a vis recent mergers or formation time [284, 295]), and future work should further address the connection of these properties with f_{NL}^{loc}. The most significant limitation of this work is that we used only the IllustrisTNG galaxy formation model - further work should of course be done to consider alternative galaxy formation models to establish the robustness of (or lack thereof) the ability to calibrate $b_\phi(b)$ using observable galaxy properties. By using IllustrisTNG galaxy formation model only and working with halo b_ϕ, we also neglect the response of galaxy halo occupation to LPNG, which has been shown to be non-negligible [291, 294, 353].

While we were finishing this manuscript, the preprint of Ref. [354] appeared presenting a two-tracer investigation of two IllustrisTNG galaxy populations selected by various secondary properties (such as halo concentration and $g - r$ color). While the focus of our work and Ref. [354] are similar, here we focus on DESI ELG and LRG galaxies in redshift space, specific future survey forecasts, and report $\sigma(f_{NL}^{\mathrm{loc}})$, rather than the improvement over single-tracer information. We consider both Ref. [354] and our work to be complementary, as Ref. [354] provides a clear and simplified explanation of the results of Ref. [275] applied to secondary galaxy properties, while also characterizing systematic effects and priors on b_ϕ, which we largely ignore here. In the other direction, we have provided a significantly more realistic treatment of how splitting by b_ϕ on a secondary parameter (in the special case of halo concentration) can be used in a multi-tracer setting for current and upcoming spectroscopic surveys.

We also developed the machine learning tools described in Section 4.3, which incorporate several observable quantities to predict b_ϕ. Our ability to constrain $f_{NL}^{\rm loc}$ is, however, limited by the dependence of b_ϕ on concentration and mass, while Ref. [354] makes use of the dependence b_ϕ on quantities such as $(g - r)$ color directly through hydrodynamical Separate Universe simulations. It would be interesting to go beyond both our work here and the work of Ref. [354] to further explore the optimal choice of observable characteristics by which to split a tracer population to obtain maximally different b_ϕ values with ML methods using different galaxy formation models.

4.5   Conclusions

Galaxy surveys contain information about the field content of the inflationary universe through the influence of local Primordial non-Gaussianity on dark matter halo formation. To extract this information, we must assume a relationship between the number density of LSS tracers and primordial fields, which is quantified by the LPNG bias parameter b_ϕ. In this work, we trained a machine learning model to connect this parameter to observable properties of DESI galaxies by way of host halo mass and concentration using simulated IllustrisTNG galaxies. We then used the learned relationship between the linear bias b_1 and LPNG bias b_ϕ for simulated galaxies to perform multi-tracer Fisher forecasts for the amplitude of LPNG $f_{NL}^{\rm loc}$ in DESI using galaxy power spectra. We also illustrated that for future LSS galaxy surveys that will target LPNG (MegaMapper, SPHEREx), there is potentially significant untapped information on $f_{NL}^{\rm loc}$ that can be extracted with the multi-tracer method when the samples can be split by host halo concentration.
We summarize our major conclusions here:

- We find that when considering simulated IllustrisTNG galaxies in a realistic setting (with observed DESI number densities, and including redshift-space distortions), the ML method can only extract part of the information about halo concentration from observable properties. This illustrates the challenge of using (unobservable) halo properties for accessing information on LPNG. We may however be limited by the small size of IllustrisTNG training data, rather than by the predictive power of observables, in which case our predictions are conservative.

- Nevertheless, our ML method extracts predictions for b_ϕ that contain enough information to allow multi-tracer Fisher forecasts of $\sigma(f_{NL}^{\rm loc})$ to greatly improve upon single-tracer forecasts. For certain sub-sample selections of DESI ELGs and LRGs, we find reductions of $\sigma(f_{NL}^{\rm loc})$ that are frequently a factor of 50%, and up to a factor of 3.

- We also provide forecasts for the "ideal" case in which b_ϕ information is perfectly known for each halo. In that case, even further reductions in $\sigma(f_{NL}^{\rm loc})$ are possible - up to a factor of 5, motivating further improvement of the methods developed here.

- We argue that future spectroscopic surveys targeting PNG (such as SPHEREx and MegaMapper) could greatly improve their constraining power on LPNG by employing a strategy similar to the one presented here, and support this argument with UMF-based forecasts for these surveys using varying Δp. More careful investigation of the potential sub-sample multi-tracer gains in LPNG information from these surveys is required.

It will be interesting to further explore to what extent the forecasts we have presented are attainable in real data analyses, and to what extent it is physically possible to obtain very large values of Δb_ϕ in cosmologies that are consistent with observed data. While here we have only explored the case of galaxy multi-tracer two-point functions; the multi-tracer bispectrum [355], and cross-correlation with a matter tracer (e.g. via CMB lensing [270]) are potential directions in which to extend the methods presented here.

Acknowledgments

4.6   Appendix A: $F_{f_{NL}^{\rm loc} f_{NL}^{\rm loc}}$ expressions

We first reproduce the expressions for the 2-tracer case, where
$\mathbf{C}^{(2)} = \frac{1}{V} \begin{pmatrix} \alpha^2 P_2 + \frac{1}{\bar{n}_1} & r\alpha P_2 \\ & P_2 + \frac{1}{\bar{n}_2} \end{pmatrix}$, and add $F_{\alpha P_2}$ to the expressions for $F_{\alpha\alpha}$ and $F_{P_2 P_2}$ of Ref. [275][7].

$$F_{\alpha\alpha} = \frac{\alpha^2 r^4 + r^2 (X_2 + 1)(X_1 - 3\alpha^2) + 2\alpha^2 (X_2 + 1)^2}{\left(\alpha^2 (-r^2 + X_2 + 1) + X_1(X_2 + 1)\right)^2}$$

$$F_{P_2 P_2} = \frac{2\alpha^2 X_1 \left(r^2(X_2 - 1) + 1\right) + \alpha^4 \left(2r^4 - 2r^2(X_2 + 2) + X_2(X_2 + 2) + 2\right) + X_1^2}{2\left(\alpha^2 P_2 (-r^2 + X_2 + 1) + P_2 X_1(X_2 + 1)\right)^2}$$

$$F_{\alpha P_2} = \frac{\alpha r^2 X_1 X_2 + \alpha^3 \left(-r^2 + X_2 + 1\right)^2}{P_2 \left(\alpha^2 (-r^2 + X_2 + 1) + X_1(X_2 + 1)\right)^2},$$

---

[7]We find a slight difference here in the expression for $F_{P_2 P_2}$ with respect to Ref. [275], though this does not affect any of the conclusions of Ref. [275] since they are all stated in the sample-variance limit, where this difference disappears.

where $X_i \equiv \frac{1}{\bar{n}_i P_2}$ quantifies the relative signal-to-noise.

In this work we used explicit symbolic expressions for the 3-tracer case, extending the expressions of Ref. [275]. The Fisher matrix element for $f_{NL}^{\rm loc}$ is given by the simple transformation:

$$F_{f_{NL}^{\rm loc} f_{NL}^{\rm loc}} = \sum_{\lambda, \lambda'} \frac{\partial \lambda}{\partial f_{NL}^{\rm loc}} F_{\lambda \lambda'} \frac{\partial \lambda'}{\partial f_{NL}^{\rm loc}} \tag{4.11}$$

where, in this work, $\lambda, \lambda' \in \{\alpha, \beta, P_2\}$. This reduces to the two-tracer expressions when β is neglected.

For the 3-tracer expressions, to make contact with the notation of Ref. [275], we can write eqn. 4.7 as

$$\mathbf{C}^{(3)} = \frac{1}{V} \begin{pmatrix} \alpha^2 P_2 + \frac{1}{\bar{n}_1} & r_{12}\alpha P_2 & r_{13}\alpha\beta P_2 \\ & P_2 + \frac{1}{\bar{n}_2} & r_{23}\beta P_2 \\ & & \beta^2 P_2 + \frac{1}{\bar{n}_3} \end{pmatrix} \tag{4.12}$$

The 3-tracer Fisher matrix element expressions for the multi-tracer parameters α, β, P_2 are then given by:

$$
\begin{aligned}
N_{\alpha\alpha} =\ & \alpha^2 r_{12}^4 (\beta^2 + X_3)^2 - 4\alpha^2\beta^2 r_{12}^3 r_{13} r_{23}(\beta^2 + X_3) \\
& + r_{12}^2(\alpha^2(\beta^4(2r_{13}^2(2r_{23}^2 + X_2 + 1) + 3(r_{23}^2 - X_2 - 1)) \\
& + \beta^2 X_3(2r_{13}^2(X_2 + 1) + 3r_{23}^2 - 6(X_2 + 1)) \\
& - 3(X_2 + 1)X_3^2) + X_1(\beta^2 + X_3)(\beta^2(-r_{23}^2 + X_2 + 1) + (X_2 + 1)X_3)) \\
& - 2\beta^2 r_{12} r_{13} r_{23}(\alpha^2(\beta^2(2r_{13}^2(X_2 + 1) + 3(r_{23}^2 - X_2 - 1)) \\
& - 3(X_2 + 1)X_3) + \beta^2 X_1(-r_{23}^2 + X_2 + 1) \\
& + X_1(X_2 + 1)X_3) + \beta^2(X_2 + 1)X_3(r_{13}^2(X_2 + 1)(X_1 - 3\alpha^2) + 4\alpha^2(-r_{23}^2 + X_2 + 1)) \\
& + \beta^4(\alpha^2 r_{13}^4(X_2 + 1)^2 - r_{13}^2(X_2 + 1)(r_{23}^2 - X_2 - 1)(X_1 - 3\alpha^2) \\
& + 2\alpha^2(-r_{23}^2 + X_2 + 1)^2) + 2\alpha^2(X_2 + 1)^2 X_3^2 \\
N_{\alpha\alpha} =\ & (\alpha^2(X_3(-r_{12}^2 + X_2 + 1) - \beta^2(r_{12}^2 - 2r_{12} r_{13} r_{23} + (r_{13}^2 - 1)X_2 + r_{13}^2 + r_{23}^2 - 1)) \\
& + \beta^2 X_1(-r_{23}^2 + X_2 + 1) + X_1(X_2 + 1)X_3)^2 \\
F_{\alpha\alpha} =\ & \frac{N_{\alpha\alpha}}{D_{\alpha\alpha}}
\end{aligned}
$$

$$
\begin{aligned}
N_{\alpha\beta} =\ & \alpha\beta(-r_{12}r_{23} + r_{13}X_2 + r_{13})(\beta^2 r_{13}(\alpha^2(r_{12}^2 - 2r_{12} r_{13} r_{23} \\
& + (r_{13}^2 - 1)X_2 + r_{13}^2 + r_{23}^2 - 1) + X_1(r_{23}^2 - X_2 - 1)) + \alpha^2 r_{13}X_3(r_{12}^2 - X_2 - 1) \\
& + X_1 X_3(-2r_{12}r_{23} + r_{13}X_2 + r_{13})) \\
D_{\alpha\beta} =\ & (\alpha^2(X_3(-r_{12}^2 + X_2 + 1) - \beta^2(r_{12}^2 - 2r_{12} r_{13} r_{23} + (r_{13}^2 - 1)X_2 + r_{13}^2 + r_{23}^2 - 1)) \\
& + \beta^2 X_1(-r_{23}^2 + X_2 + 1) + X_1(X_2 + 1)X_3)^2 \\
F_{\alpha\beta} =\ & \frac{N_{\alpha\beta}}{D_{\alpha\beta}}
\end{aligned}
$$

$$
\begin{aligned}
N_{P_2\alpha} =\ & \alpha(\alpha^2 r_{12}^4(\beta^2 + X_3)^2 - 4\alpha^2\beta^2 r_{12}^3 r_{13}r_{23}(\beta^2 + X_3) + r_{12}^2(2\alpha^2(\beta^4(r_{13}^2(2r_{23}^2 + X_2 + 1) \\
& + r_{23}^2 - X_2 - 1) \\
& + \beta^2 X_3((r_{13}^2 - 2)X_2 + r_{13}^2 + r_{23}^2 - 2) - (X_2 + 1)X_3^2) + X_1(\beta^2 r_{23}^2 X_3 + X_2(\beta^2 + X_3)^2)) \\
& - 2\beta^2 r_{12}r_{13}r_{23}(2\alpha^2(\beta^2((r_{13}^2 - 1)X_2 + r_{13}^2 + r_{23}^2 - 1) - (X_2 + 1)X_3) \\
& + X_1(\beta^2 X_2 + 2X_2 X_3 + X_3)) \\
& + \beta^4(r_{13}^2 r_{23}^2 X_1 X_2 + \alpha^2(r_{13}^2(X_2 + 1) + r_{23}^2 - X_2 - 1)^2) \\
& + \beta^2(X_2 + 1)X_3(r_{13}^2(X_2 + 1)(X_1 - 2\alpha^2) \\
& + 2\alpha^2(-r_{23}^2 + X_2 + 1)) + \alpha^2(X_2 + 1)^2 X_3^2) \\
D_{P_2\alpha} =\ & P_2(\alpha^2(X_3(-r_{12}^2 + X_2 + 1) - \beta^2(r_{12}^2 - 2r_{12}r_{13}r_{23} + (r_{13}^2 - 1)X_2 + r_{13}^2 + r_{23}^2 - 1)) \\
& + \beta^2 X_1(-r_{23}^2 + X_2 + 1) + X_1(X_2 + 1)X_3)^2 \\
F_{P_2\alpha} =\ & \frac{N_{P_2\alpha}}{D_{P_2\alpha}}
\end{aligned}
$$

$$
\begin{aligned}
N_{P_2 P_2} =\ & 2\alpha^2 X_1(\beta^4(2(r_{23}^2 - 1)(r_{12}^2 - 2r_{12}r_{13}r_{23} + r_{13}^2 + r_{23}^2 - 1) \\
& + 2(r_{13}^2 - 1)(r_{23}^2 - 1)X_2 - (r_{13}^2 - 1)X_2^2) + \beta^2 X_3(2r_{12}^2(r_{23}^2 + X_2 - 1) \\
& - 6r_{12}r_{13}r_{23}X_2 + r_{13}^2 X_2(X_2 + 2) + 2r_{23}^2(X_2 - 1) + 2) + X_3^2(r_{12}^2(X_2 - 1) + 1)) \\
& + \alpha^4(\beta^4(4(r_{13}^2 - 1)X_2(r_{12}^2 - 2r_{12}r_{13}r_{23} + r_{13}^2 + r_{23}^2 - 1) \\
& + 3(r_{12}^2 - 2r_{12}r_{13}r_{23} + r_{13}^2 + r_{23}^2 - 1)^2 \\
& + 2(r_{13}^2 - 1)^2 X_2^2) + 2\beta^2 X_3(2(r_{12}^2 - 1)(r_{12}^2 - 2r_{12}r_{13}r_{23} + r_{13}^2 + r_{23}^2 - 1) \\
& + 2(r_{12}^2 - 1)(r_{13}^2 - 1)X_2 \\
& - (r_{13}^2 - 1)X_2^2) + X_3^2(2r_{12}^4 - 2r_{12}^2(X_2 + 2) + X_2(X_2 + 2) + 2)) \\
& + X_1^2(2\beta^2 X_3(r_{23}^2(X_2 - 1) + 1) + \beta^4(2r_{23}^4 - 2r_{23}^2(X_2 + 2) + X_2(X_2 + 2) + 2) + X_3^2) \\
D_{P_2 P_2} =\ & 2P_2^2(\alpha^2(X_3(-r_{12}^2 + X_2 + 1) - \beta^2(r_{12}^2 - 2r_{12}r_{13}r_{23} + (r_{13}^2 - 1)X_2 + r_{13}^2 + r_{23}^2 - 1)) \\
& + \beta^2 X_1 \\
& (-r_{23}^2 + X_2 + 1) + X_1(X_2 + 1)X_3)^2 \\
F_{P_2 P_2} =\ & \frac{N_{P_2 P_2}}{D_{P_2 P_2}}
\end{aligned}
$$

We do not write the expressions for $F_{\beta P_2}$ or $F_{\beta\beta}$ since these are the same as $F_{\alpha P_2}, F_{\alpha\alpha}$ with interchanged arguments $\beta \longleftrightarrow \alpha$.

The Jacobian factors for the transformation in eqn. 4.11 for the parameters α and β are:

$$
\frac{\partial \alpha}{\partial f_{NL}^{\text{loc}}} = \alpha\left(\frac{b_{\phi,1}}{b_1} - \frac{b_{\phi,2}}{b_2}\right)\mathcal{M}^{-1} \tag{4.13}
$$

$$
\frac{\partial P_2}{\partial f_{NL}^{\text{loc}}} = 2b_{\phi,2}\mathcal{M}^{-1}\frac{P_2}{b_2} \tag{4.14}
$$

where we dropped the k-dependence of $\mathcal{M}$ and P_2 to be consistent with the previous expressions, and again the β expression is analogous to the α expression.

In the sample variance limit, at leading order in X_i, and assuming $b_1 = b_2 = b_3$, the ratio of the $f_{NL}^{\rm loc}$ Fisher information is:

$$
\frac{F_{f_{NL}^{\rm loc}}^{(3)}}{F_{f_{NL}^{\rm loc} f_{NL}^{\rm loc}}^{(2)}} = \frac{2\left((b_{\phi,1} - b_{\phi,2})^2 + (b_{\phi,2} - b_{\phi,3})^2 + (b_{\phi,1} - b_{\phi,3})^2\right)}{3(b_{\phi,1} - b_{\phi,2})^2}
$$
$$
- \frac{2X_2\left(43b_{\phi,1}^2 + 59b_{\phi,1}b_{\phi,2} - 145b_{\phi,1}b_{\phi,3} - 59b_{\phi,2}^2 + 59b_{\phi,2}b_{\phi,3} + 43b_{\phi,3}^2\right)}{27(b_{\phi,1} - b_{\phi,2})^2}
\tag{4.15}
$$

$$
\frac{F_{f_{NL}^{\rm loc} f_{NL}^{\rm loc}}^{(3,P)}}{F_{f_{NL}^{\rm loc} f_{NL}^{\rm loc}}^{(2,P)}} = \frac{F_{f_{NL}^{\rm loc}}^{(3)}}{F_{f_{NL}^{\rm loc} f_{NL}^{\rm loc}}^{(2)}} - \frac{4b_{\phi,2}X_2}{9(b_{\phi,1} - b_{\phi,2})^4}[18b_{\phi,1}^3 + b_{\phi,1}^2(33b_{\phi,2} - 86b_{\phi,3})
$$
$$
+ 4b_{\phi,1}\left(-4b_{\phi,2}^2 + 4b_{\phi,2}b_{\phi,3} + 13b_{\phi,3}^2\right)
$$
$$
+ 17b_{\phi,2}^2(b_{\phi,2} - 2b_{\phi,3})]
\tag{4.16}
$$

when considering only relative amplitude parameters, and when including P_2, respectively. These expressions illustrate that these contributions both approach the first line of eqn. 4.15 as $X_2 \to 0$, but can be significantly different otherwise.

Similarly, when considering the ratio of the Fisher information when using P_2 as a parameter or not for two tracers, we have:

$$
\frac{F_{f_{NL}^{\rm loc}}^{(2,P)}}{F_{f_{NL}^{\rm loc} f_{NL}^{\rm loc}}^{(2)}} = 1 + X_2\frac{52b_{\phi,1}b_{\phi,2}}{3(b_{\phi,1} - b_{\phi,2})^2}.
\tag{4.17}
$$

Again, we see that in the sample variance limit the Fisher information is equivalent, but outside of the case where $X_2 \to 0$ the exact values of b_ϕ in each sample can drive the ratio above or below one (e.g. in the latter case if the values of $b_{\phi,1}, b_{\phi,2}$ have different signs).

We now make a few brief comments about these expressions, which may be useful for the interested reader of the main text. To be consistent with the notation of Ref. [275] we have taken the two-tracer case to have LPNG biases $b_{\phi,1}$ and $b_{\phi,2}$, so care should be taken to interpret these expressions appropriately (and not to simply associate $b_{\phi,1}$ with b_ϕ^- or $b_{\phi,3}$ with b_ϕ^+). In Section 4.4, for future surveys we considered a symmetric Δp for the upper and lower tertiles, though in general this will not be the case. Eqn. 4.15 also indicates that the exact departure from this symmetry may affect whether a 2-tracer of 3-tracer forecast is desirable. We also note that, in general, both the linear bias b of the sample and the number density $\bar{n}$ will spoil these statements as we leave the sample variance limit and b changes over the splits.

Chapter 5

Deterministic Langevin Optimization

This chapter was originally posted as: James M. Sullivan, and Uroš Seljak. "Deterministic Langevin Unconstrained Optimization with Normalizing Flows". Submitted to Journal of Global Optimization, DOI:10.48550/arXiv.2310.00745, arXiv:2310.00745 [cs.LG]

In this chapter, We introduce a global, gradient-free surrogate optimization strategy for expensive black-box functions inspired by the Fokker-Planck and Langevin equations. These can be written as an optimization problem where the objective is the target function to maximize minus the logarithm of the current density of evaluated samples. This objective balances exploitation of the target objective with exploration of low-density regions. The method, Deterministic Langevin Optimization (DLO), relies on a Normalizing Flow density estimate to perform active learning and select proposal points for evaluation. This strategy differs qualitatively from the widely-used acquisition functions employed by Bayesian Optimization methods, and can accommodate a range of surrogate choices. We demonstrate superior or competitive progress toward objective optima on standard synthetic test functions, as well as on non-convex and multi-modal posteriors of moderate dimension. On real-world objectives, such as scientific and neural network hyperparameter optimization, DLO is competitive with state-of-the-art baselines.

5.1   Introduction

Gradient-free global black-box optimization is perhaps the most widely-shared task among scientific and engineering applications. Such problems are often non-convex, multi-modal, and rugged - presenting a serious challenge for any local optimization strategy [356]. While there are a host of well-known global methods with which to attack this problem (e.g. population-based methods [357], gradient-based solvers with finite difference approximations [358], evolutionary algorithms [359], simplex-based methods [360]), alternatives must be considered in the limit of very expensive objective function evaluations. By comparison, active learning (AL) [361] refers to a class of methods suited to problems where objective evaluation is expensive and requires the next point in the input parameter space to be very

carefully chosen. A popular form of active learning is Bayesian Optimization (BO), which has demonstrated strong performance in the limit of expensive objective evaluation (which translates into a small function call budget for fixed wall-clock time) [362, 363, 364, 365, 366, 367, 368]. BO is a probabilistic method that iterates over steps consisting of two components: 1. fitting a surrogate model to the objective function and 2. selecting new points at which to evaluate the expensive objective (via an "acquisition function") - the goal is to strike a balance between exploitation and exploration. The details of BO then boil down to what kind of surrogate model and acquisition function (AF) are chosen.

Gaussian Processes (GPs) are almost always the core of BO algorithms due to their analytic form and interpolation properties [369, 370]. GPs provide both a surrogate model, through the GP mean $\mu(\theta)$, as well as, crucially, an estimate of uncertainty at unseen points, through the GP error $\sigma(\theta)$ [371]. The uncertainty estimate is governed by the form of the GP kernel and its hyperparameters, which effectively provide a smoothness prior on the objective function, and is necessary to prioritize exploration to avoid getting stuck in local optima. GP uncertainties, while theoretically compelling, are not the only tool for exploration in the context of active learning. Due partially to the poor scaling of linear algebra operations necessary for GP fitting and evaluation, and partially due to the related GP issues in high-dimensional problems [372], BO-like AL methods have been proposed with alternative surrogates. Such alternatives include methods that use neural networks (NN) [373, 374], random forests [375], or radial basis functions [376] as surrogate models. However, to produce an exploration strategy in the absence of GP uncertainty, such models employ strategies such as model ensembling to estimate uncertainty. Ensembling is often empirically effective, but can lead to overconfident uncertainty estimates and can be computationally expensive for deep models like NNs [377, 378, 379, 380].

In this contribution we argue that there is another way to estimate the uncertainty of the surrogate model, which is to use density estimation: if the local density of the sampling points is relatively high in the parameter space of interest, then the surrogate uncertainty will be lower than other locations in this space, and vice versa. We propose to use normalizing flow (NF) density estimation of evaluated points to estimate this uncertainty. This allows us to develop an exploration strategy using the NF density estimate by constructing a novel acquisition function, which is theoretically motivated by deterministic Langevin dynamics. We demonstrate that this acquisition function, when combined with a local-global hybrid strategy similar to that recently proposed for the trust-region BO method of [381], meets or exceeds the performance of this method as well as that of widely-used evolutionary and finite-difference gradient algorithms on standard test functions and objectives relevant for probabilistic inference in science and engineering.

5.2 Deterministic Langevin Optimization

We will phrase all discussion in terms of maximization of a scalar objective $f(\theta)$, input parameters $\theta \in \mathbb{R}^d$, and in the context of the problem $\theta^* \in \arg_\theta \max f(\theta)$. When the goal of

optimization is minimization we change the sign of f.

Theoretical Motivation

The overdamped Langevin equation is a stochastic differential equation describing particle motion in an external potential and subject to a random force with zero mean,

$$\frac{d\theta}{dt} = v = \beta \nabla_\theta f(\theta) + \eta, \tag{5.1}$$

where $\langle \eta(t) \rangle = 0$ and $\langle \eta_i(t) \eta_j(t') \rangle = 2\delta_{ij}\delta(t - t')$ where $\langle \rangle$ denotes the expectation taken wrt. the distribution of η, v is the velocity of the particles with position θ (with $v, \theta \in \mathbb{R}^d$), and β is the inverse temperature. We set the diffusion coefficient to unity.

The Langevin equation can be viewed as a particle implementation of the evolution of the (unnormalized) particle probability density $q(\theta_t)$, which is governed by the deterministic Fokker-Planck equation, a continuity equation for the density,

$$\frac{dq(\theta_t)}{dt} + \nabla_\theta \cdot J = 0, \tag{5.2}$$

$$J = q(\theta_t)\nabla_\theta[\beta f(\theta_t) + V(\theta_t)] \equiv q(\theta_t)v. \tag{5.3}$$

Here θ_t is the particle position at a some time t and we defined $V(\theta_t) = -\ln q(\theta_t)$ and expressed the current J as density times velocity, where the two terms in the probability current J correspond to the two velocity terms in the Langevin equation. When we reach a stationary distribution where $\frac{dq(\theta_t)}{dt} = 0$, the corresponding density is given by $q(\theta) \propto \exp(\beta f(\theta))$. Thus the solution of the Fokker-Planck equation, combined with a suitable temperature annealing, where we start with $\beta = 0$ and end with $\beta \to \infty$, will lead us to the solution of the optimization problem.

In practice, solving the Fokker-Planck equation in high dimensions is difficult [382]. Instead, if we replace the stochastic velocity in the Langevin equation 5.1 with the deterministic velocity in equation 5.3, we obtain the *deterministic Langevin equation* [383, 384], which in discretized form is

$$\theta_{t+1} = \theta_t + v\epsilon = \theta_t + \nabla_\theta[\beta f(\theta_t) + V(\theta_t)]\epsilon, \tag{5.4}$$

where ϵ is the step size and V_t is the negative logarithm of the density defined by the positions $\{\theta_n\}$ of all previous evaluations. The first term in brackets on the right-hand side of this equation moves the particles in the direction of the target peak $\beta f(\theta)$, while the second term moves them in the direction of low density as defined by the points that have been previously evaluated. Equation 5.4 can thus be interpreted heuristically as a gradient-based maximization of the objective $\beta f(\theta) + V_t(\theta)$, or

$$\theta_{t+1}^\star \in \arg\max_\theta[\beta f(\theta) + V(\theta)] = \arg\max_\theta \ln \frac{\exp(\beta f(\theta))}{q_t(\theta)}. \tag{5.5}$$

This equation makes manifest the exploration versus exploitation nature of the deterministic Langevin objective: we can either move into the region of the highest $\exp(\beta f(\theta))$ (exploitation), or we can move into the region of the lowest $q_t(\theta)$, which we may not have explored yet (exploration). It is not our intention to mimic the diffusion process of the Fokker-Planck or Deterministic Langevin equations, but to motivate the placement of the next sampling point using the largest discrepancy between the current density and target density. This target objective is time index t-dependent: as we explore a region by evaluating particles in it, its density increases and this reduces the objective at that position. Furthermore, we can also adjust the temperature annealing such that the target objective $\exp(\beta f(\theta))$ is shallow initially, and peaked at the end.

Surrogate and DLO acquisition function

To make optimization of the objective of equation 5.5 into a working algorithm we still need several ingredients. First, we need to be able to propose values of θ based on current knowledge of $f(\theta)$ via an acquisition function, which requires a surrogate model. For the surrogate model $s(\theta)$ of the target $f(\theta)$ in the acquisition function, we deploy Gaussian Processes (GPs) in the main results of this work, as the GP is the standard surrogate choice in Bayesian Optimization algorithms. However, we do not take advantage of the GP error estimate, so other surrogates such as neural networks (NN) [373, 374], random forests [375], or radial basis functions [376] could also be used. We provide a limited discussion extending our results, which are presented for only moderate dimension, to higher dimension using simple neural network surrogates in Appendix 5.9. The other ingredient of the acquisition function is the uncertainty estimate, $q(\theta)$, which we return to in Section 5.2.

We incorporate the surrogate into the DLO objective equation 5.5 by employing the following acquisition function

$$\text{DLO}(\theta; \beta) \equiv s(\theta; \beta) - \ln q_t(\theta) = \ln \frac{e^{s(\theta; \beta)}}{q_t(\theta)}. \tag{5.6}$$

Here the surrogate is fitted to the annealed target as $s(\theta; \beta) = \mathcal{GP}(\{\theta\}, \{\beta \times f(\theta)\})$ - i.e., the GP takes as arguments the current "data" given by $\{\theta\}, \{\beta \times f(\theta)\}$ pairs. In our default implementation of DLO, $s(\theta)$ is a GP mean, while q_t is an NF fitted to the sample density after t sample evaluations.

Normalizing Flows

The main difficulty in solving equation 5.5 is in evaluating the instantaneous density term $q_t(\theta)$. Here we address this by feeding all previously evaluated points $\theta_1, ...\theta_t$ into a normalizing flow (NF) to evaluate the density of these samples.

Normalizing flows are deep generative models that provide a map from a simple "base" distribution, such as a uniform or standard normal distribution, to a desired target distribution. Normalizing flows provide a powerful framework for density estimation and sampling
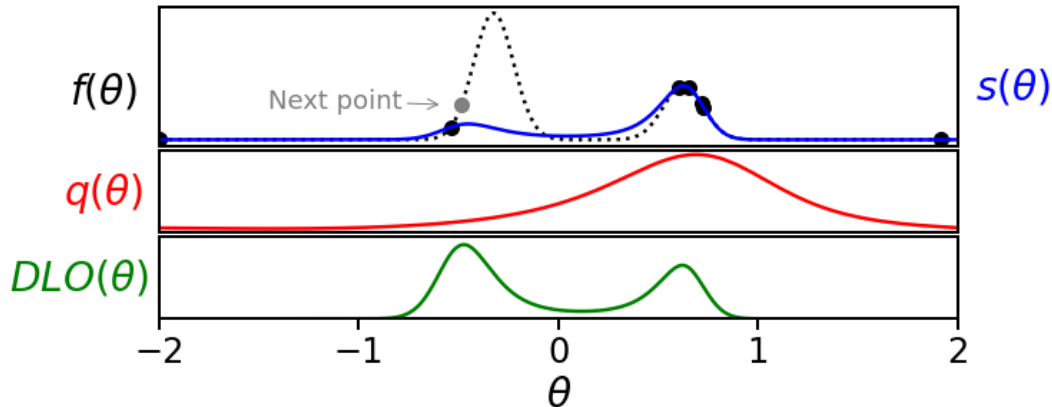
Figure 5.1: A schematic depiction of the Deterministic Langevin Optimization algorithm. The key elements of the algorithm are shown from top to bottom at an early (β_i) stage of the algorithm (with 7 calls) applied to a mixture of Gaussians. *Top:* The surrogate model $s(\theta, \beta_i)$ (solid) begins to fit the target objective (black dotted). *Center:* The density estimate $q(\theta)$ gives a smooth density estimate. *Bottom:* The acquisition function DLO (equation 5.6) determines where the next point will be selected (gray point). At this stage of DLO, the algorithm has determined that the smaller peak is sufficiently explored (density q is high) and now turns to explore the second local peak of surrogate containing the true maximum. The next iteration following the one in the illustration identifies the true maximum at $\theta = -0.325$.

[385, 386, 387, 138]. These models map the d-dimensional data θ to d-dimensional latent variables z through a sequence of invertible transformations $\Psi = \Psi_1 \circ \Psi_2 \circ ... \circ \Psi_L$, such that $z = \Psi(\theta)$ and z is mapped to a base distribution $\pi(z)$, which we choose to be a standard Normal distribution $N(0, \boldsymbol{I})$. The probability density of data θ can be evaluated using the change of variables formula:

$$q(\theta) \equiv e^{-V(\theta)} = \pi(\Psi(\theta))|\det(\frac{\partial \Psi(\theta)}{\partial \theta})|$$

$$= \pi(\Psi(\theta))\prod_{l=1}^{L}|\det(\frac{\partial \Psi_l(\theta)}{\partial \theta})|. \tag{5.7}$$

The Jacobian determinant of each transform $J_l = |\det(\partial \Psi_l(\theta)/\partial \theta)|$ must be easy to compute in order to evaluate the density, and the transformation Ψ_l should be easy to invert for efficient sampling. In this paper we use the Sliced Iterative Normalizing Flow (SINF) [138] NF algorithm. SINF scales well to high dimensions and shows good performance on small training sets while using a low number of hyperparameters. The details of SINF are discussed in the Appendix 5.7.

Local Exploration

To optimize the DLO objective of equation 5.6 we could use gradient-based optimization, since both the GP and NF are differentiable. However, this local strategy will be susceptible to getting stuck in local extrema. Instead, we use a local exploration strategy that is common in BO, drawing a fixed number of proposal samples N_{sample} around the points of higher DLO objective. We choose $N_{\mathrm{sample}} = 100d$.

To generate half of the proposal domain points, we adopt a local exploration strategy similar to [381], generating domain points within a hyperrectangle (a Gaussian sphere also works well) characterized by a length scale R that grows and shrinks with iteration. As in [381], R is initialized to a prescribed starting value and grows and shrinks by a logarithmic step dR depending on how many iterations have passed since $\Delta f_i > 0$.

The bijective map of the normalizing flow permits efficient sampling of the target distribution. We make use of this capability by sampling the other half of proposal samples at each iteration from the latent space of the density estimation normalizing flow $q_t(\theta)$. This is especially helpful for sharply-peaked high-dimensional objectives, as we avoid considering proposals in large regions of essentially zero objective value (even within our local proposal volume). We draw half of the proposal samples from a Gaussian sphere of radius R in the latent space of the normalizing flow centered on θ^* after it has been mapped into latent space. Once the proposal points are chosen we evaluate their DLO objective of equation 5.6 and choose the highest value. For more details on the effect of removing local exploration from DLO when applied to test objectives, see Appendix 5.10.

Simulated Annealing

We employ simulated annealing [388] to reduce the scale of variation of our target objective early on in the DLO iteration procedure. The surrogate model fitted to the annealed posterior is denoted $s(\theta, \beta)$, where β is the annealing parameter. It is well-known that simulated annealing is beneficial for optimizing rugged or multi-modal objectives, and is also valuable for population-based Bayesian sampling strategies applied to such challenging functions [389]. We want to design an annealing scheme, where we vary β between initial and final values β_0 and β_{max}, such that we achieve a good simultaneous exploration and exploitation strategy. We set the inverse temperature (β) simulated annealing schedule using N_β logarithmically-spaced steps in β, where N_β is determined by the objective function call budget N and batch size B (with B supplied by the user based on computational constraints), $N_\beta = (N - N_I)/B$, where N_I is the number of initial samples.

In our experiments we observe that there is no single value of β_{max} that gives the best performance but on many examples $\beta_{\mathrm{max}} = 100$ and $\beta_{\mathrm{max}} \to \infty$ perform well. The latter ignores NF: this is a good strategy when the objective function is isotropic and broad. In this case, NF density estimation does not help much with the optimization strategy. In other situations, specifically for posterior objectives where large fractions of the input space have very low density, NF exploration is essential for optimal results. We note that using only

$\beta_{\max} = 100$ still leads to competitive performance. As a result, we will choose between these two values based on their performance, starting with an annealing scheme that combines them with a 50-50 split, which can be adjusted during the annealing if needed according to their performance.

We also need to set the initial annealing level β_0. We select β_0 such that the difference between the largest and smallest value of $f(\theta)$ on $N_I = 2d$ initial samples is less than 15. This ensures sufficiently smooth GP interpolation of initial samples such that early on, if one initial function value dominates the rest by several orders of magnitude, the smaller values are not ignored if they are non-zero (this is especially beneficial for very narrow high-dimensional objectives). Choosing $\beta_0 \neq 1$ performs a similar function as standardizing the objective function output (as is done, e.g., in [381] using an affine transformation).

Sometimes the β_0 condition is already satisfied for $\beta_{\max} = 100$. In this case we do not anneal at all. For all of the synthetic objectives in Section 5.3, the range of objective function variation is small enough that annealing is not needed. However, for posterior objectives and hyper-parameter optimization (HPO) the annealing schedule is essential for good performance. In practice, there is a broad range of β_0 values that give the same performance. For more details on the effect of removing the simulated annealing step from DLO when applied to test objectives, see Appendix 5.10.

Algorithm

The high-level Deterministic Langevin Optimization algorithm is presented in Algorithm 1. We rescale the domain of θ to the product of d unit intervals. Each update $\theta_t \to \theta_{t+1}$ contains a batch of B samples.

Algorithm 1 Deterministic Langevin Optimization

1: Evaluate $f(\theta_1), .., f(\theta_{N_I})$ at N_I initial points; select initial annealing level β_0, rescale the input θ domain to $[0, 1]^d$, decide number of AF proposal points per iteration N_{sample}, and the batch size B.

2: Assign a call budget N, fix the hyperparameters N_β, R, dR.

3: **for** $i < N_\beta$ **do**

4:    Estimate the normalizing flow density $q_i(\theta)$ from $\theta_1, .., \theta_t$.

5:    Fit the surrogate $s_i(\theta, \beta_i))$ from $f(\theta_1), ..f(\theta_t)$ to annealed objective values.

6:    Create proposal samples in $[0, 1]^d$ *and* in the latent space of q_t drawing from Gaussian spheres of radius R around the highest $\text{DLO}(\theta_j), j = 1...t$.

7:    Locally maximize the acquisition function $\text{DLO}(\theta)$ from N_{sample} proposal draws to obtain the next batch of $\theta_{t+1}, .., \theta_{t+B}$ to evaluate.

8:    Evaluate $f(\theta_{t+1}), ..f(\theta_{t+B})$ and update β_i.

9: **end for**

5.3 Numerical Experiments

We test DLO on three types of test objectives in this section. We first show performance on several standard gradient-free optimization test functions, before turning to posterior objectives relevant for applying DLO to Bayesian posterior optimization relevant for inference in science and engineering, and finally demonstrate that DLO is competitive on several applied objectives, including Hyperparameter Optimization (HPO). We also compare our new acquisition function eqn. 5.6 to several alternatives that hew closer to the literature.

We compare DLO to `TuRBO`, which, for general-purpose single-objective gradient-free optimization, is an extremely popular highly-performing BO method. The local optimization aspect of our method is based on a trust region strategy [381] - so this comparison is especially relevant for illustrating the improvement furnished by our modified acquisition function. We use the `TuRBO` hyperparameters and numerical setups as detailed in [381] unless otherwise stated. We further compare to differential evolution (DiffEv) [357] and the covariance matrix adaptation evolution strategy (CMA-ES) [390], which give an indication of how population-based and evolutionary strategies perform in the limit of a low number of function evaluations, and to the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm with finite difference gradients. We also include a random search baseline. For all experiments we consider only sequential evaluation (batch size of 1) function evaluations for DLO unless otherwise noted. Details of the numerical experiments are provided in Appendix 5.6.

Synthetic Objectives

We consider two common synthetic test functions in the gradient-free optimization literature in 10 dimensions - the Ackley function and Rastrigin function. Both functions are characterized by the presence of many local optima, with the Ackley function having a global optimum that is narrow and much higher than the local minima, while the Rastrigin function has a shallow global curvature with more pronounced local optima. We use domains of $[-5, 10]^{10}$ and $[-5.12, 5.12]^d$ for the Ackley and Rastrigin functions, respectively. These functions are usually given as minimization objectives. Since our method is formulated with posterior inference applications in mind, we transform these synthetic objectives into maximization problems as $g_{\min} \to g_{\max} \equiv \log\left(g(\theta^*) - g(\theta)\right)$.

Results on these functions are presented in Figure. 5.2, where we see that DLO outperforms the other methods in the low function call budget regime (over 30 optimization runs). On both objectives, DLO sees an extra jump in performance early on in the optimization runs with respect to the other methods, which can be attributed to our modified acquisition function. The evolutionary methods and BFGS, meanwhile, are not competitive with DLO and `TuRBO` on these functions.
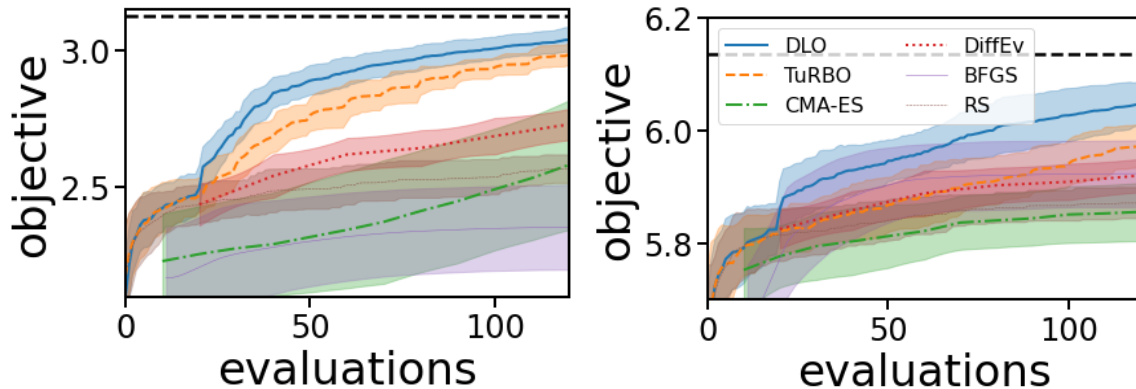
Figure 5.2: Performance of DLO relative to other methods on converging to the optimum on $10 - d$ synthetic objectives. *Left:* Ackley function. *Right:* Rastrigin function. Solid lines show the mean over all optimization runs, and shaded regions show the standard deviations. The dark dashed line is the true value of the objective function at the global optimum.

Posterior Objectives

A key application of AL methods is to expensive scientific inference problems [35]. Such problems (e.g. physical simulations [142]) may require model evaluations that take minutes, hours, or days of wall-clock time on multiple nodes of high-performance computing facilities. Efficient parameter estimation with as few samples as possible is thus of paramount importance, and MAP optimization is faster than Monte Carlo Markov Chains, which are often prohibitively expensive. With this wide class of applications in mind, we consider several 10-dimensional test posterior objectives of the type that frequently arise in scientific parametric inference applications.

A correlated Gaussian distribution is the simplest non-trivial posterior routinely encountered in statistical inference. We make this problem challenging by enforcing a high condition number (ratio of maximum to minimum eigenvalues) of 200 in the test posterior. Figure 5.3 shows that DLO (and `TuRBO`) acheive the best performance, and are consistent with each other, though DLO has a slight edge early-on. BFGS also attains decent performance, but with a large variance between runs.

Posteriors that resemble Gaussian mixtures arise frequently in scientific inference (e.g. in inferring gravitational wave merger parameters [391]). We consider a widely-separated double Gaussian posterior with thin peaks that has a false maximum (one peak is higher than the other, and contains 70% of the probability mass). While the difficulty of this example produces a large variance between runs (especially for the evolutionary algorithms), Figure 5.3 shows that DLO typically finds a point near the true maximum first, followed by `TuRBO`, and eventually the evolutionary algorithms, while later-evaluation performance of DLO and `TuRBO` is similar. For this example we note that the gradient-based BFGS typically converges first - but this is to the wrong peak 9 times out of 30, and unlike the
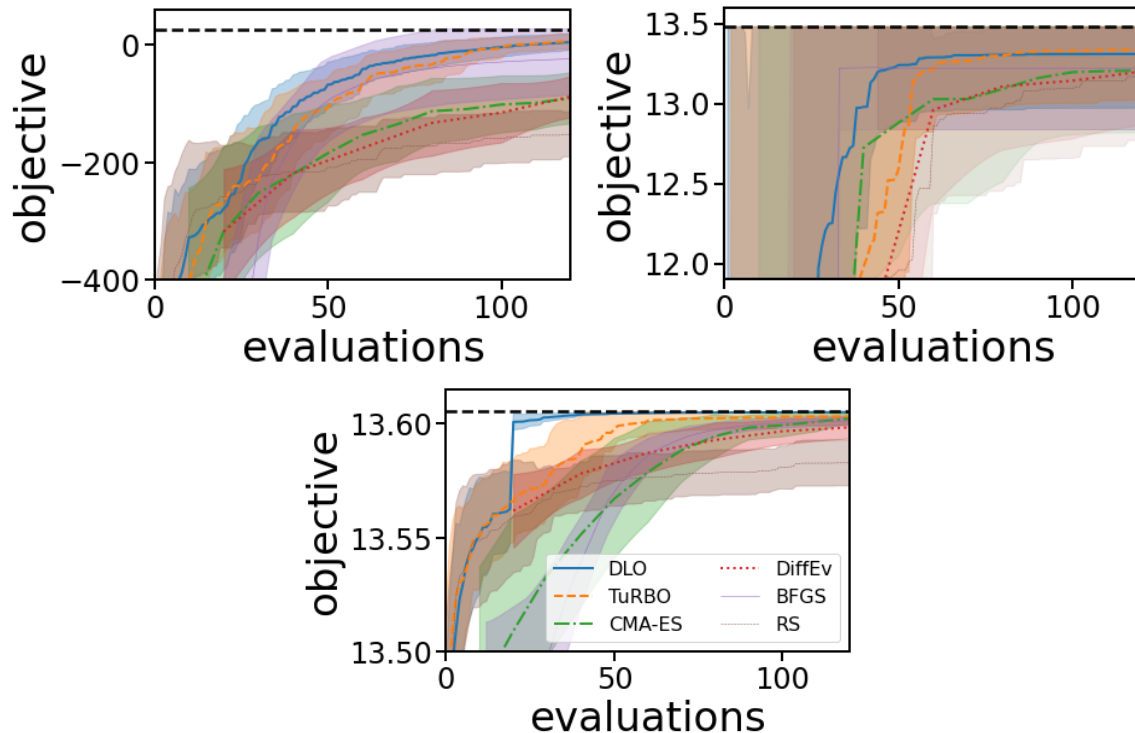
Figure 5.3: Performance of DLO relative to other methods on converging to the optimum on $10d$ posterior objectives. *Left*: Correlated Gaussian with condition number 200. *Right*: Widely-separated double Gaussian mixture. *Bottom*: Rosenbrock posterior. Solid lines show the mean over all optimization runs, and shaded regions show the standard deviations.

other algorithms, has no chance at improving after this point.

We also consider the non-convex Rosenbrock function as a posterior (after suitable shifting to ensure a positive objective), which is a test frequently considered by Bayesian sampling methods [392] due to thin curving posteriors that frequently arise in scientific applications. We find that DLO attains the best performance on this objective by far, jumping up to the optimum before `TuRBO` and all other methods (including BFGS).

Applied Objectives

"Cosmological Constants"

The "Cosmological Constants" problem is the optimization of a $12d$ posterior objective relevant for the analysis of cosmological parameters from Sloan Digital Sky Survey Luminous Red Galaxy clustering data [393]. The dominant cost of the posterior evaluation is the solution of a large system of stiff differential equations - the Einstein-Boltzmann equations. For this problem we fix the neutrino mass to zero (making the problem $11d$), which greatly
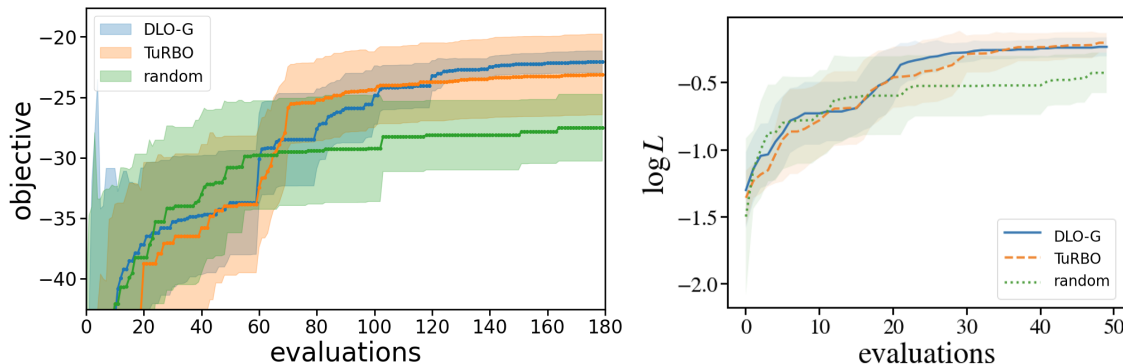
Figure 5.4: Performance of DLO compared to two baselines (`TuRBO` and random search) on converging to the optimum on several applied objectives. *Left*: "Cosmological Constants" posterior *Right*: Neural network (Multi-layer Perceptron) HPO (blue line corresponds to DLO). Solid lines show the mean over all optimization runs, and shaded regions show the standard deviations.

reduces the time-to-solution of the Einstein-Boltzmann solver, but leave all other parameters free. Figure 5.4 illustrates that DLO outperforms TuRBO (averaged over 15 runs).

Machine Learning Model Hyperparameter Optimization

BO-style AL methods are well-suited to hyperparameter optimization (HPO) for machine learning models, and have begun to gain traction over random/grid search among practitioners [394, 395, 396, 397] since they outperform such simple methods [398, 399, 400, 401]. In particular, the `TuRBO` baseline and its competitor-supplied variants performed well in the 2020 Bayesian Black-Box optimization challenge for several simple models on publicly available (scikit learn) datasets. We consider an example drawn from this challenge, in particular we use the `bayesmark`[1] experiment for multi-layer perceptron SGD applied to the `scikit-learn` [402] "iris" dataset with the (negative) log likelihood objective, in Figure 5.4. This Figure shows that DLO is competitive with `TuRBO` (over 15 replications) on this HPO problem, and gives results that are consistent with `TuRBO` in this small-budget setting.

Modifying the acquisition function

We perform an acquisition function ablation study to characterize the extent to which the NF density estimate equation 5.6 is responsible for the strong performance of DLO shown in Sections 5.3-5.3. Figure 5.5 shows progress toward the optimum of DLO when applied to the $10d$ Rastrigin function for several common BO acquisition functions: Expected Improvement (EI) [403], Upper Confidence Bound (UCB) [404], and Thompson Sampling (TS) [405, 406].

---

[1]`https://bayesmark.readthedocs.io/`

These AFs all incorporate the GP uncertainty estimate to guide the extent to which to prioritize exploration, and, for reference, are given by:

$$\text{EI}(\theta) = (\mu(\theta) - f^*(\theta))\Phi\left(\frac{(\mu(\theta) - f^*(\theta))}{\sigma(\theta)}\right)$$
$$+ \sigma(\theta)\phi\left(\frac{(\mu(\theta) - f^*(\theta))}{\sigma(\theta)}\right) \tag{5.8}$$

$$\text{UCB}(\theta) = \mu(\theta) + \beta_{\text{UCB}}\sigma(\theta), \tag{5.9}$$

$$\text{TS}(\theta) = f \sim \mathcal{GP}(\mu(\theta), \sigma(\theta)). \tag{5.10}$$

and we set $\beta_{\text{UCB}} = 1$. Figure 5.5 shows that DLO (with $\beta_{\text{max}} = 100$) has a distinct edge over the other acquisition functions, especially very early on in the optimization runs.

5.4 Related Work

Gradient-free optimization is one of the most active areas of global optimization and there is a correspondingly vast literature covering many methods. We review some aspects of these methods that are most relevant to DLO.

Local+global methods: The `TuRBO` strategy [381] has led to many follow-up methods (e.g. [407]) and applications - including in constrained problems [408], BO with axis-aligned features [409], and multi-objective problems [410]. Basin hopping has long been used by the chemistry community for optimizing multi-modal functions [411], and other global BO methods with local BFGS searches have recently been developed for atomic structures [412]. The DLO global strategy differs from these, while it can still take advantage of the many possible local methods proposed in the literature.

Alternative uncertainty estimates: Several studies have explored supplying a different surrogate model in place of a GP, which must come with a different uncertainty estimate. Alternatives that have been explored include random forests (`SMAC` [375]), neural networks (`DNGO` [373], `BOHAMIANN` [374]). While such surrogates differ from the standard GP, these methods have retained versions of ([413]) Expected Improvement (EI) as their acquisition functions by estimating the variance a GP would have provided from these surrogate models (i.e. by looking at variation over RF trees or via a Bayesian last-layer treatment). `NOMU` [378] incorporates a more complex NN uncertainty estimate by training two dependent NNs and using a carefully constructed (non-EI) loss function. DLO replaces uncertainty quantification with density estimation: the density of samples in some region is lower, it is expected that the interpolation quality of the surrogate will be lower, which we can interpret as higher surrogate uncertainty.

Modified acquisition functions: Ref. [414] proposed an acquisition function that places increased weight on BO exploration with the goal of Bayesian posterior approximation. This was employed in a BO framework applied to scientific simulators [415, 416], but was restricted to relatively low-dimensional examples. Other recent work [417, 418, 418, 419] has

explored non-myopic acquisition functions, which make the choice of which point to evaluate based on predictions of future evaluations. These predictions are made by (approximately) integrating over future GP predictions, and therefore employ GP uncertainty. Entropy-based AF methods have also been proposed as an alternative to GP-based uncertainty for AL [420, 421, 422]. DLO differs from these methods, as it works directly with the density of the sampled points, rather than the density estimated by entropy-based methods. The `BORE` method [423, 424] employs density-ratio estimation as part of its acquisition strategy in the context of formulating EI as a classification task, which is distinct from our evaluated sample density estimation. DLO AF differs from all of these methods, since it uses NF density estimation as part of AF.
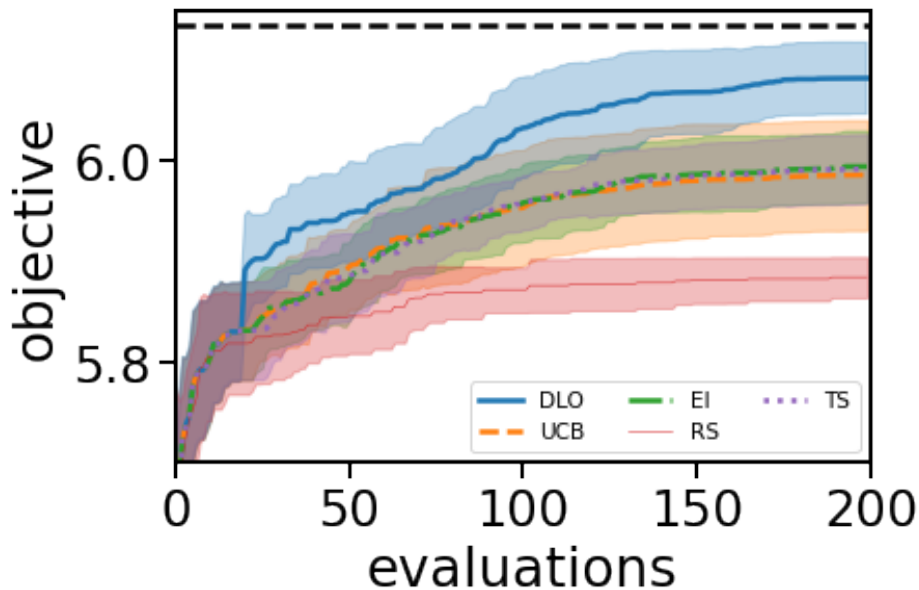
Figure 5.5: Performance of the DLO algorithm for several choices of standard GP acquisition function on the 10d Rastrigin function. DLO shows a clear improvement over the other acquisition strategies. Solid lines show the mean over all optimization runs, and shaded regions show the standard deviations.

5.5 Conclusions

Probabilistic surrogate models are a powerful component of gradient-free black-box global optimization. While variants of Gaussian Process-based Bayesian Optimization have demonstrated enormous success on this class of tasks, using a Gaussian Process-based acquisition strategy is not the only competitive choice. We have demonstrated that Normalizing Flow density estimation provides an exploration strategy that is competitive, and frequently su-

perior, to that provided by GP-based methods in the context of BO-like active learning optimization when incorporated into the Deterministic Langevin acquisition function we propose in this paper. We demonstrated this through the optimization of several standard optimization test functions, common posterior objectives arising in Bayesian inference, and real-world application objectives, such as cosmological inference and ML hyperparameter optimization.

The proposed method is not without limitations. One is the choice of the hyperparameter $\beta_{\max}$ and the annealing schedule. We do not expect that one choice will fit all problems, a consequence of the no free-lunch theorem of global optimization: in some problems we can use a greedy strategy and progress quickly towards the peak without much penalty, in others we need more exploration to find all the existing peaks. This can be controlled by the choice of $\beta_{\max}$. In this paper we compare the performance between two different values of $\beta_{\max}$, but there may be better ways to do this.

Another limitation is that the wall-clock time of DLO is currently limited by the $\mathcal{O}(n^3)$ scaling of evaluation of the default Exact GP we employ in `gpytorch` [425]. This may be alleviated through GPU-based methods for accelerating GP kernel operations when many samples are present [426] (for the experiments presented here were run on a single 2.8 GHz Quad-Core Intel Core i7 CPU), or through the use of a more efficient surrogate (we briefly explore this in Appendix 5.9).

5.6 Appendix A: Setup of Numerical Experiments

For all experimental results, we use $N_0 = 2d$ initialization points for each numerical experiment for DLO and `TuRBO` sampled from a Latin hypercube (except when the implementation of the algorithms, such as for BFGS ad CMA-ES, only permit a single initial value, which is generated from a uniform distribution on the input space in that case). We use a total function call budget of $N_0 + 10d$ for each problem, except in Section 5.3, where we use twice this budget. We use the `scipy` implementation of differential evolution and the CMA-ES implementation of [427]. Though this comparison was already performed in [381], we also include the `scipy` [428] implementation of the Broyden-Fletcher-Goldfarb-Shanno (`BFGS`) algorithm [358] to demonstrate the challenges gradient-based methods face in the low-budget regime on the objectives we consider. If BFGS converges before the function call budget is reached, we set the rest of the values in the experiment to the converged value.

The Ackley and Rastrigin functions are given (in 10 dimensions) by

$$F(\theta) = 20 \left(1 - \exp\left[-\frac{1}{5}\sqrt{\frac{1}{10}\sum_i^{10}\theta_i^2} \right] \right)$$
$$+ e \left(1 - \exp\left[\frac{1}{10}\sum_i^{10}\cos\left(2\pi\theta_i\right) \right] \right),$$
$$F(\theta) = 100 + \sum_i^{10}\left(\theta_i^2 - 10\cos(2\pi\theta_i) \right),$$

respectively. Their domains are $[-5, 10]^{10}$ and $[-5.12, 5.12]^{10}$.

The posterior objectives we consider are the correlated Gaussian posterior, which is simply a Gaussian distribution with mean $\mu = \frac{1}{5}\mathbf{u}$, where $\mathbf{u}$ is the $10d$ vector whose entries are all ones, and covariance C such that for the singular values λ_i of C, $\frac{\lambda_1}{\lambda_{10}} = 200$, with $\lambda_1 = 0.09$.

The double Gaussian posterior is the mixture

$$F(\theta) = 0.3p_N(\theta; 0.625, 0.1)$$
$$+ 0.7p_N(\theta; -0.325, 0.1)$$

where $p_N(\theta; \mu, \sigma)$ is the multi-variate Gaussian probability density. We restrict to the domain $[-2, 2]^{10}$ for both the correlated and double Gaussian posteriors. The Rosenbrock posterior is given by

$$F(\theta) = \sum_{i=0}^{9}(1 - \theta_i)^2 + 100\left(\theta_{i+1} - \theta_i^2 \right)^2$$

on the domain $[-5, 5]^{10}$.

The "Cosmological constants" example is described in the main text - but involves the maximization of an $11d$ log likelihood for the galaxy power spectrum - a compressed statistic of galaxy survey data - to obtain a point estimate of cosmological parameters. These parameters include the matter density, expansion rate, and various numerical factors that govern the differential equation solver that serves as part of the model called by the likelihood.

The HPO example we consider uses the "iris" dataset and the `bayesmark` (`sklearn`) implementation of a Multi-layer Perceptron which has parameters that are optimized through SGD, and uses the (inverted) negative log likelihood as the objective. The hyperparameters for this model are the number of hidden layers, the learning rate and its initialization, the batch size, the tolerance, the momentum, a regularization parameter (`alpha`), and learning rate scaling.

5.7 Appendix B: DLO Hyperparameters

In this section we discuss the DLO hyperparameters.

The GP surrogate model is implemented in `pytorch` [429] and uses the `gpytorch` [425] ExactGP implementation. Following [381], we use a GP with Matérn-5/2 covariance, but force the GP noise to be smaller (between $10^{-6} - 10^{-4}$), and do not standardize the objective function values. We optimize over the GP hyperparameters using 50 steps of `ADAM` [430].

For the normalizing flow we employ for density estimation (SINF), we use 5 iterations, no regularization and up to 8 sliced Wasserstein directions (the default prescription of [**dai2021sliced**]). The most important hyperparameter of SINF is a pre-factor bw that modifies a Scott's rule prescription for the SINF density estimation pre-factor (see [**dai2021sliced**] equation 26). We demonstrate that variation of this hyperparameter has very little effect on our results in Figure 5.6. In this study $X = 1$, though the bw dependence does not greatly depend on this fact. Clearly, very small bw (less than 1) leads to somewhat reduced performance, while if bw is "large enough" there is no change in performance. Intuitively this makes sense - a too small bw means that the density estimating flow q will be close to zero except very close to the evaluated objective points, where it will be very high. This will mean proposals in the support of the density (which is tightly hugging the evaluated points) will have very low AF value. While a lower AF value in these regions is somewhat desirable (otherwise there will never be any explicit exploration - exploration would only be possible through the implicit error on $s(\theta; \beta)$), if the already-evaluated points are high in the objective, this will slow down convergence to the optimum. A larger bw effectively regularizes the value of q to avoid the above scenario by softening the penalty near already-evaluated points. We fix $bw = 1.0$ to avoid reducing performance - this is also the default value in SINF.
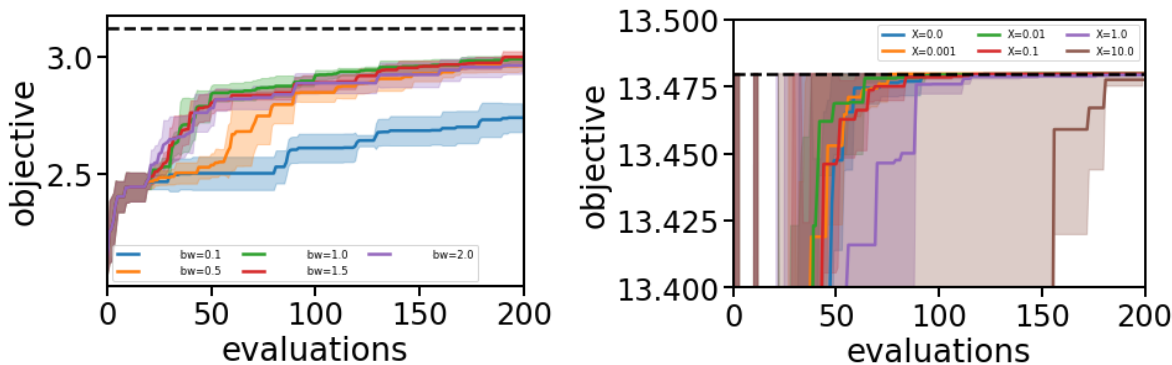
Figure 5.6: *Left:* Dependence of progress toward the optimum on the NF bandwidth factor for the GP-based version of DLO with mean and standard error estimated over 5 realizations. *Right:* Dependence of progress toward the optimum on the AF coefficient X for the GP-based version of DLO with mean and standard error estimated over 5 realizations.

There is no a-priori way to know how much exploration should be weighted against exploitation. While the standard acquistion functions EI and TS have no explicit hyper-parameters in their functional form (that is, outside of those used in constructing the GP), UCB (which we note performs best in our tests, c.f. Fig. 5.5) contains the hyperparameter β_{UCB}. While theoretical estimates have been provided for β_{UCB} (e.g. in [404]), in practice, the same authors choose a value that works better empirically. In our setting this is controlled by the value $\beta_{\max}$. We explore the effect of changing $\beta_{\max}$ on the double Gaussian example in Figure 5.6. Clearly too-large values of X lead to over exploration and slow convergence, while small values show improved convergence. We select $X = 0.01$ as our fiducial value to avoid over-exploring. The DLO proposal volume evolution is similar to that of `TuRBO` (which is drawn from the trust region literature). The volume starts at $R = 1$ and decreases by dR if we do not improve (within a tolerance of 5×10^{-6}) on the best value f^* after two iterations. After a successful improvement on f^*, the volume increases by the same factor to consider a larger proposal volume. Rather than using Automatic Relevance Determination to fit an independent GP kernel lengthscale parameter in each coordinate dimension, when not proposing in latent space, we scale the proposal volume along the coordinate axes by the gradient direction unit vector.

5.8 Appendix C: Batch size

We show the effect of batch size on our method on the synthetic Ackley test function in Figure 5.7 (here using $X = 1$ for all proposals). A too-large batch leads to somewhat reduced performance, as would be expected due to each batch having less information to use when selecting the next point with the acquisition function.
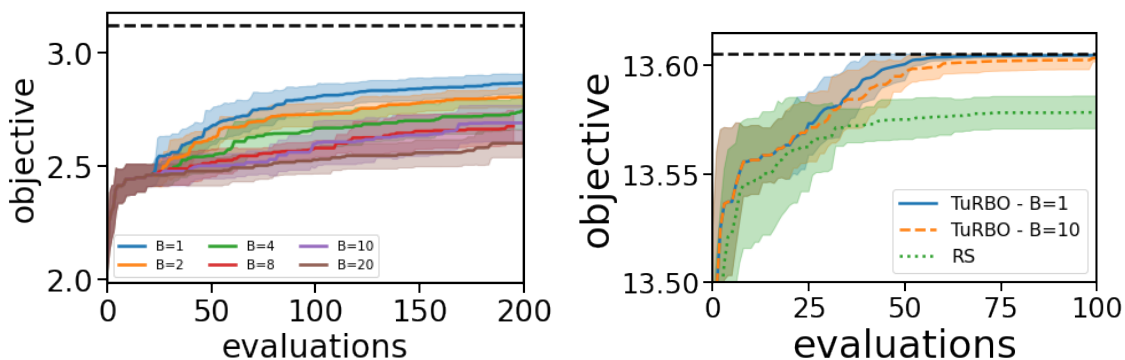
Figure 5.7: *Left:* Dependence of progress toward the optimum for the GP-based version of DLO with mean and standard error estimated over 10 realizations. *Right:* Optimization runs on the $10d$ Rosenbrock function, as in Figure 5.3, but changing the `TuRBO` batch size from 10 (as provided by those authors for $10d$ test functions) to 1. We also show a random search baseline.

Figure 5.7 indicates the change in performance of `TuRBO` when the batch size is reduced to 1 from the values prescribed in [381]. Following the results presented in Figure 7 of [381], we see that the performance improves with reduced batch size, but only slightly.

5.9   Appendix D: Replacing the surrogate model

In the main text, we presented an implementation of the DLO algorithm that employed a GP for the surrogate model and an NF, SINF, for the density estimate. However, as the number of dimensions of the problem at hand increases, more data points are required to obtain an accurate surrogate model. Since the cost of the GP surrogate scales as the number of data points cubed, alternative surrogates may be preferred in high dimensions.

In this spirit, we consider a simple fully-connected neural network surrogate as a replacement for the GP mean surrogate. In Figure 5.8 we use a FCN implemented in `pytorch` with 2 hidden layers of width 100 trained with ADAM [430] and Tanh activation as the surrogate and compare it to the progress toward the optimum on the Rastrigin function in $10d$. We find slightly reduced performance but a generally comparable result, indicating deep network surrogates can perform well as replacement for GPs. We have not expended significant time or effort on finding a well-performing architecture, so results can likely be specifically improved upon those shown here, but this is out of the scope of this work, as is further exploration of higher dimensional problems. However, already it is empirically clear that the NN surrogate is much more scalable than the GP. Table 5.1 makes this cost scaling explicit, as while the neural network is always cheaper to fit and evaluate as a surrogate than a GP in the context of DLO, the scaling of the GP cost is significantly worse with dimension for the Ackley objective (here both methods obtain similar performance in terms of the final objective value).
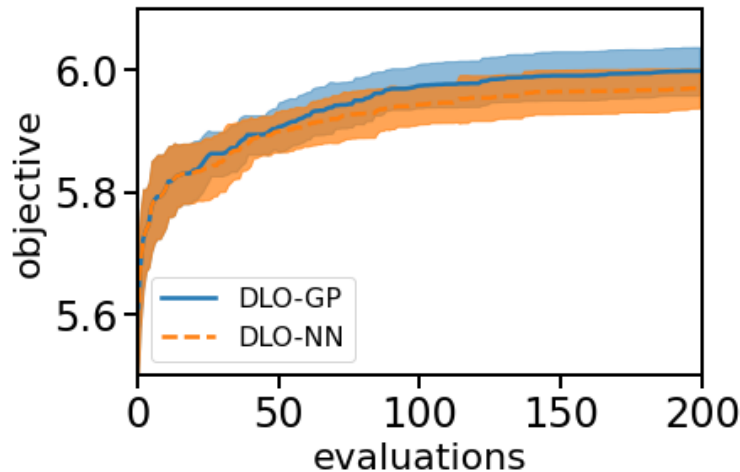
Figure 5.8: Progress toward objective optimum on the 10-d Rastrigin test problem for both a simple neural network surrogate and the default choice of a Gaussian Process surrogate. Here we average over 15 realizations and provide the related confidence band.

d	2	5	10	20	50
Evaluation:					
DLO-GP	0.02	0.03	0.07	0.15	6.07
DLO-NN	0.02	0.02	0.04	0.10	0.53
Fitting:					
DLO-GP	0.17	0.26	0.44	1.71	46.65
DLO-NN	0.03	0.04	0.06	0.20	2.84

Table 5.1: Surrogate *evaluation* (top rows) and *fitting* (bottom rows) timings (in seconds) for both Gaussian Process (GP) and fully-connected neural network (NN) surrogates at the last iteration of optimization of the Ackely function for several choices of dimension d.

5.10   Appendix E: Annealing and local search ablation

To demonstrate that both the annealing and suggesting candidate points from a local search volume (rather than from a global search) are necessary for the performance shown in the main text, here we perform a simple ablation study and remove these two components of the DLO algorithm. Figure 5.9 illustrates that for the 10-dimensional objectives, performance suffers when annealing and the local bounding box for choosing candidate points are omitted.

In particular, for the Ackley function, taking out the local volume search gives significantly worse average performance (in terms of objective value obtained), while removing annealing has comparatively little effect. For the correlated Gaussian posterior, the situation is reversed - removing the local volume search seems to have little effect, while the omission of annealing causes more variation between runs and worse average empirical convergence toward the optimum. The differences here are not surprising, as the two targets furnish very different optimization problems. For one, the Gaussian target is convex, and has an enormous dynamic range in objective $(\log p(\theta))$ value, which is dramatically reduced by the temperature annealing. Meanwhile, the Ackley function has comparatively little dynamic range, and has many local minima, so it is possible for the acquisition strategy to overfavor far-away local minima if they have not yet been explored when there is no local candidate search volume.
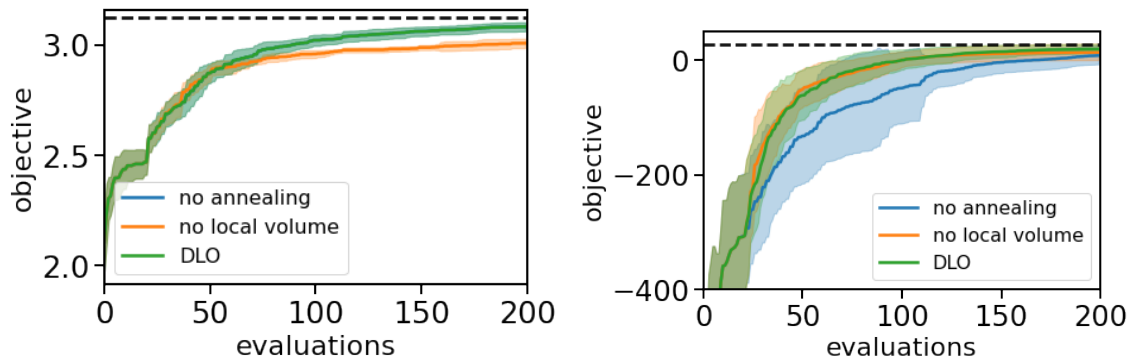
Figure 5.9: The effect of removing temperature annealing and the local volume search from the DLO algorithm on performance in 10 dimensions for the Ackley function (*left*) and the correlated Gaussian (*right*) objectives (averaged over 15 realizations).

5.11   Appendix F: Further synthetic test functions

We also show results on two additional test functions, the $8-$dimensional Zakharov function on $[-5, 10]^8$ and the $12-$dimensional Styblinski-Tang function, on $[-5, 5]^{12}$. The Zakharov function is valley-shaped, and the Styblinski-Tang is only slightly multi-modal, significantly different from the Rastrigin and Ackley functions. DLO is shown in both its GP surrogate mode ("DLO-G") and fully-connected network mode ("DLO-N"). We see that DLO struggles somewhat with the Zakharov function but handily outperforms the other methods for the Styblinski-Tang
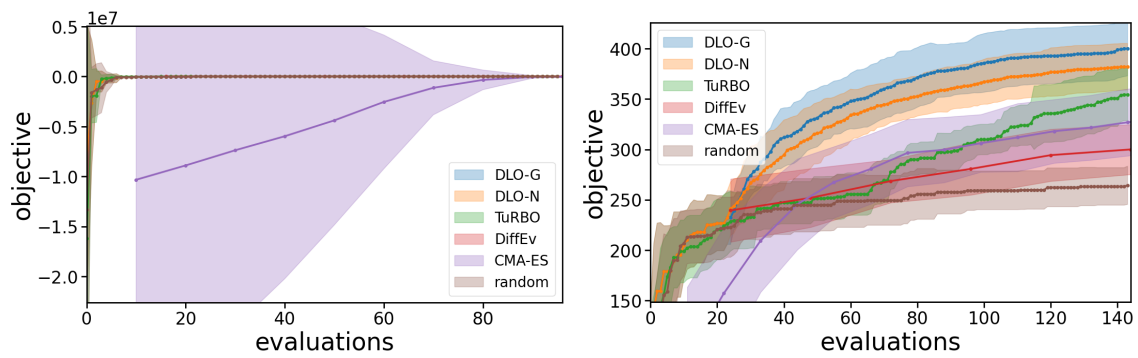
Figure 5.10: Similar to Fig. 5.2, but for the $8 - d$ Zakharov (*left*) and the $12 - d$ Styblinski-Tang (*right*) objectives (averaged over 15 realizations, with standard deviation as the shaded area).

Bibliography

[1] Particle Data Group et al. "Review of Particle Physics". In: *Progress of Theoretical and Experimental Physics* 2022.8 (Aug. 2022), p. 083C01. ISSN: 2050-3911. DOI: 10.1093/ptep/ptac097. eprint: https://academic.oup.com/ptep/article-pdf/2022/8/083C01/49175539/ptac097.pdf. URL: https://doi.org/10.1093/ptep/ptac097.

[2] Julien Lesgourgues and Sergio Pastor. "Massive neutrinos and cosmology". In: Phys. Rep. 429.6 (July 2006), pp. 307–379. DOI: 10.1016/j.physrep.2006.04.001. arXiv: astro-ph/0603494 [astro-ph].

[3] 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].

[4] Daniel Baumann. "TASI Lectures on Inflation". In: *arXiv e-prints*, arXiv:0907.5424 (July 2009), arXiv:0907.5424. DOI: 10.48550/arXiv.0907.5424. arXiv: 0907.5424 [hep-th].

[5] Viatcheslav Mukhanov. *Physical Foundations of Cosmology.* 2005. DOI: 10.2277/0521563984.

[6] Lev Kofman, Andrei Linde, and Alexei A. Starobinsky. "Reheating after inflation". In: Phys. Rev. Lett. 73.24 (Dec. 1994), pp. 3195–3198. DOI: 10.1103/PhysRevLett.73.3195. arXiv: hep-th/9405187 [hep-th].

[7] Rouzbeh Allahverdi et al. "Reheating in Inflationary Cosmology: Theory and Applications". In: *Annual Review of Nuclear and Particle Science* 60 (Nov. 2010), pp. 27–51. DOI: 10.1146/annurev.nucl.012809.104511. arXiv: 1001.2600 [hep-th].

[8] Rouzbeh Allahverdi et al. "The First Three Seconds: a Review of Possible Expansion Histories of the Early Universe". In: *The Open Journal of Astrophysics* 4.1, 1 (Jan. 2021), p. 1. DOI: 10.21105/astro.2006.16182. arXiv: 2006.16182 [astro-ph.CO].

[9] Mustafa A. Amin et al. "Nonperturbative dynamics of reheating after inflation: A review". In: *International Journal of Modern Physics D* 24.1, 1530003 (Dec. 2015), p. 1530003. DOI: 10.1142/S0218271815300037. arXiv: 1410.3808 [hep-ph].

[10] Alan H. Guth and David I. Kaiser. "Inflationary Cosmology: Exploring the Universe from the Smallest to the Largest Scales". In: *Science* 307.5711 (Feb. 2005), pp. 884–890. DOI: 10.1126/science.1107483. arXiv: astro-ph/0502328 [astro-ph].

[11] 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].

[12] Scott Dodelson and Fabian Schmidt. *Modern Cosmology*. 2020. DOI: 10.1016/C2017-0-01943-2.

[13] Chung-Pei Ma and Edmund Bertschinger. "Cosmological Perturbation Theory in the Synchronous and Conformal Newtonian Gauges". In: ApJ 455 (Dec. 1995), p. 7. DOI: 10.1086/176550. arXiv: astro-ph/9506072 [astro-ph].

[14] Katrin Heitmann et al. "The Last Journey. I. An Extreme-scale Simulation on the Mira Supercomputer". In: ApJS 252.2, 19 (Feb. 2021), p. 19. DOI: 10.3847/1538-4365/abcc67. arXiv: 2006.01697 [astro-ph.CO].

[15] F. Bernardeau et al. "Large-scale structure of the Universe and cosmological perturbation theory". In: Phys. Rep. 367.1-3 (Sept. 2002), pp. 1–248. DOI: 10.1016/S0370-1573(02)00135-7. arXiv: astro-ph/0112551 [astro-ph].

[16] Y. B. Zel'Dovich. "Reprint of 1970A&A.....5...84Z. Gravitational instability: an approximate theory for large density perturbations." In: A&A 500 (Mar. 1970), pp. 13–18.

[17] Martin White. "The Zel'dovich approximation". In: MNRAS 439.4 (Apr. 2014), pp. 3630–3640. DOI: 10.1093/mnras/stu209. arXiv: 1401.5466 [astro-ph.CO].

[18] Takahiko Matsubara. "Resumming cosmological perturbations via the Lagrangian picture: One-loop results in real space and in redshift space". In: Phys. Rev. D 77.6, 063530 (Mar. 2008), p. 063530. DOI: 10.1103/PhysRevD.77.063530. arXiv: 0711.2521 [astro-ph].

[19] Mikhail M. Ivanov. "Effective Field Theory for Large Scale Structure". In: *arXiv e-prints*, arXiv:2212.08488 (Dec. 2022), arXiv:2212.08488. DOI: 10.48550/arXiv.2212.08488. arXiv: 2212.08488 [astro-ph.CO].

[20] Raul E. Angulo and Oliver Hahn. "Large-scale dark matter simulations". In: *Living Reviews in Computational Astrophysics* 8.1, 1 (Dec. 2022), p. 1. DOI: 10.1007/s41115-021-00013-z. arXiv: 2112.05165 [astro-ph.CO].

[21] Vincent Desjacques, Donghui Jeong, and Fabian Schmidt. "Large-scale galaxy bias". In: Phys. Rep. 733 (Feb. 2018), pp. 1–193. DOI: 10.1016/j.physrep.2017.12.002. arXiv: 1611.09787 [astro-ph.CO].

[22] Volker Springel, Carlos S. Frenk, and Simon D. M. White. "The large-scale structure of the Universe". In: Nature 440.7088 (Apr. 2006), pp. 1137–1144. DOI: 10.1038/nature04805. arXiv: astro-ph/0604561 [astro-ph].

[23] James Binney and Scott Tremaine. *Galactic Dynamics: Second Edition*. 2008.

[24] Houjun Mo, Frank C. van den Bosch, and Simon White. *Galaxy Formation and Evolution*. 2010.

[25] Asantha Cooray and Ravi Sheth. "Halo models of large scale structure". In: Phys. Rep. 372.1 (Dec. 2002), pp. 1–129. DOI: 10.1016/S0370-1573(02)00276-4. arXiv: astro-ph/0206508 [astro-ph].

[26] Risa H. Wechsler and Jeremy L. Tinker. "The Connection Between Galaxies and Their Dark Matter Halos". In: ARA&A 56 (Sept. 2018), pp. 435–487. DOI: 10.1146/annurev-astro-081817-051756. arXiv: 1804.03097 [astro-ph.GA].

[27] Planck Collaboration et al. "Planck 2018 results. I. Overview and the cosmological legacy of Planck". In: A&A 641, A1 (Sept. 2020), A1. DOI: 10.1051/0004-6361/201833880. arXiv: 1807.06205 [astro-ph.CO].

[28] Planck Collaboration et al. "Planck 2015 results. XIII. Cosmological parameters". In: A&A 594, A13 (Sept. 2016), A13. DOI: 10.1051/0004-6361/201525830. arXiv: 1502.01589 [astro-ph.CO].

[29] Dillon Brout et al. "The Pantheon+ Analysis: Cosmological Constraints". In: ApJ 938.2, 110 (Oct. 2022), p. 110. DOI: 10.3847/1538-4357/ac8e04. arXiv: 2202.04077 [astro-ph.CO].

[30] David Rubin et al. "Union Through UNITY: Cosmology with 2,000 SNe Using a Unified Bayesian Framework". In: *arXiv e-prints*, arXiv:2311.12098 (Nov. 2023), arXiv:2311.12098. DOI: 10.48550/arXiv.2311.12098. arXiv: 2311.12098 [astro-ph.CO].

[31] DES Collaboration et al. "The Dark Energy Survey: Cosmology Results With ~1500 New High-redshift Type Ia Supernovae Using The Full 5-year Dataset". In: *arXiv e-prints*, arXiv:2401.02929 (Jan. 2024), arXiv:2401.02929. DOI: 10.48550/arXiv.2401.02929. arXiv: 2401.02929 [astro-ph.CO].

[32] Shadab Alam et al. "Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Cosmological implications from two decades of spectroscopic surveys at the Apache Point Observatory". In: Phys. Rev. D 103.8, 083533 (Apr. 2021), p. 083533. DOI: 10.1103/PhysRevD.103.083533. arXiv: 2007.08991 [astro-ph.CO].

[33] Andrew Gelman et al. *Bayesian Data Analysis*. 2014.

[34] Steve Brooks et al. *Handbook of Markov Chain Monte Carlo*. English (US). Publisher Copyright: © 2011 by Taylor and Francis Group, LLC. All rights reserved. CRC Press, May 2011. ISBN: 9781420079418.

[35] Kyle Cranmer, Johann Brehmer, and Gilles Louppe. "The frontier of simulation-based inference". In: *arXiv e-prints*, arXiv:1911.01429 (Nov. 2019), arXiv:1911.01429. arXiv: 1911.01429 [stat.ML].

[36] M. Huertas-Company and F. Lanusse. "The Dawes Review 10: The impact of deep learning for the analysis of galaxy surveys". In: PASA 40, e001 (Jan. 2023), e001. DOI: 10.1017/pasa.2022.55. arXiv: 2210.01813 [astro-ph.IM].

[37] Uroš Seljak and Zvonimir Vlah. "Halo Zel'dovich model and perturbation theory: Dark matter power spectrum and correlation function". In: Phys. Rev. D 91.12, 123516 (June 2015), p. 123516. DOI: 10.1103/PhysRevD.91.123516. arXiv: 1501.07512 [astro-ph.CO].

[38] Shadab Alam et al. "The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological analysis of the DR12 galaxy sample". In: MNRAS 470.3 (Sept. 2017), pp. 2617–2652. DOI: 10.1093/mnras/stx721. arXiv: 1607.03155 [astro-ph.CO].

[39] T. M. C. Abbott et al. "Dark Energy Survey year 1 results: Cosmological constraints from galaxy clustering and weak lensing". In: Phys. Rev. D 98.4, 043526 (Aug. 2018), p. 043526. DOI: 10.1103/PhysRevD.98.043526. arXiv: 1708.01530 [astro-ph.CO].

[40] Vincent Desjacques, Donghui Jeong, and Fabian Schmidt. "Large-scale galaxy bias". In: Phys. Rep. 733 (Feb. 2018), pp. 1–193. DOI: 10.1016/j.physrep.2017.12.002. arXiv: 1611.09787 [astro-ph.CO].

[41] 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, 82 (Sept. 2012), p. 82. DOI: 10.1007/JHEP09(2012)082. arXiv: 1206.2926 [astro-ph.CO].

[42] Daniel Baumann et al. "Cosmological non-linearities as an effective fluid". In: J. Cosmology Astropart. Phys. 2012.7, 051 (July 2012), p. 051. DOI: 10.1088/1475-7516/2012/07/051. arXiv: 1004.2488 [astro-ph.CO].

[43] Zvonimir Vlah, Martin White, and Alejandro Aviles. "A Lagrangian effective field theory". In: J. Cosmology Astropart. Phys. 2015.9, 014 (Sept. 2015), p. 014. DOI: 10.1088/1475-7516/2015/09/014. arXiv: 1506.05264 [astro-ph.CO].

[44] Rafael A. Porto, Leonardo Senatore, and Matias Zaldarriaga. "The Lagrangian-space Effective Field Theory of large scale structures". In: J. Cosmology Astropart. Phys. 2014.5, 022 (May 2014), p. 022. DOI: 10.1088/1475-7516/2014/05/022. arXiv: 1311.2168 [astro-ph.CO].

[45] Zvonimir Vlah et al. "Perturbation theory, effective field theory, and oscillations in the power spectrum". In: J. Cosmology Astropart. Phys. 2016.3, 057 (Mar. 2016), p. 057. DOI: 10.1088/1475-7516/2016/03/057. arXiv: 1509.02120 [astro-ph.CO].

[46] Mikhail M. Ivanov, Marko Simonović, and Matias Zaldarriaga. "Cosmological parameters from the BOSS galaxy power spectrum". In: J. Cosmology Astropart. Phys. 2020.5, 042 (May 2020), p. 042. DOI: 10.1088/1475-7516/2020/05/042. arXiv: 1909.05277 [astro-ph.CO].

[47] Guido d'Amico et al. "The cosmological analysis of the SDSS/BOSS data from the Effective Field Theory of Large-Scale Structure". In: J. Cosmology Astropart. Phys. 2020.5, 005 (May 2020), p. 005. DOI: 10.1088/1475-7516/2020/05/005. arXiv: 1909.05271 [astro-ph.CO].

[48] Thomas Colas et al. "Efficient cosmological analysis of the SDSS/BOSS data from the Effective Field Theory of Large-Scale Structure". In: J. Cosmology Astropart. Phys. 2020.6, 001 (June 2020), p. 001. DOI: 10.1088/1475-7516/2020/06/001. arXiv: 1909.07951 [astro-ph.CO].

[49] Matthew McQuinn and Martin White. "Cosmological perturbation theory in 1+1 dimensions". In: J. Cosmology Astropart. Phys. 2016.1, 043 (Jan. 2016), p. 043. DOI: 10.1088/1475-7516/2016/01/043. arXiv: 1502.07389 [astro-ph.CO].

[50] Enrico Pajer and Drian van der Woude. "Divergence of perturbation theory in large scale structures". In: J. Cosmology Astropart. Phys. 2018.5, 039 (May 2018), p. 039. DOI: 10.1088/1475-7516/2018/05/039. arXiv: 1710.01736 [astro-ph.CO].

[51] Uroš Seljak. "Analytic model for galaxy and dark matter clustering". In: MNRAS 318.1 (Oct. 2000), pp. 203–213. DOI: 10.1046/j.1365-8711.2000.03715.x. arXiv: astro-ph/0001493 [astro-ph].

[52] J. A. Peacock and R. E. Smith. "Halo occupation numbers and galaxy bias". In: MNRAS 318.4 (Nov. 2000), pp. 1144–1156. DOI: 10.1046/j.1365-8711.2000.03779.x. arXiv: astro-ph/0005010 [astro-ph].

[53] Chung-Pei Ma and J. N. Fry. "Deriving the Nonlinear Cosmological Power Spectrum and Bispectrum from Analytic Dark Matter Halo Profiles and Mass Functions". In: ApJ 543.2 (Nov. 2000), pp. 503–513. DOI: 10.1086/317146. arXiv: astro-ph/0003343 [astro-ph].

[54] Benjamin D. Wibking et al. "Emulating galaxy clustering and galaxy-galaxy lensing into the deeply non-linear regime: methodology, information, and forecasts". In: MNRAS 484.1 (Mar. 2019), pp. 989–1006. DOI: 10.1093/mnras/sty2258. arXiv: 1709.07099 [astro-ph.CO].

[55] Fabian Schmidt. "Towards a self-consistent halo model for the nonlinear large-scale structure". In: Phys. Rev. D 93.6, 063512 (Mar. 2016), p. 063512. DOI: 10.1103/PhysRevD.93.063512. arXiv: 1511.02231 [astro-ph.CO].

[56] Lehman H. Garrison et al. "The Abacus Cosmos: A Suite of Cosmological N-body Simulations". In: ApJS 236.2, 43 (June 2018), p. 43. DOI: 10.3847/1538-4365/aabfd3. arXiv: 1712.05768 [astro-ph.CO].

[57] Douglas Potter, Joachim Stadel, and Romain Teyssier. "PKDGRAV3: beyond trillion particle cosmological simulations for the next era of galaxy surveys". In: *Computational Astrophysics and Cosmology* 4.1, 2 (May 2017), p. 2. DOI: `10.1186/s40668-017-0021-1`. arXiv: `1609.08621 [astro-ph.IM]`.

[58] Katrin Heitmann et al. "The Coyote Universe. II. Cosmological Models and Precision Emulation of the Nonlinear Matter Power Spectrum". In: ApJ 705.1 (Nov. 2009), pp. 156–174. DOI: `10.1088/0004-637X/705/1/156`. arXiv: `0902.0429 [astro-ph.CO]`.

[59] Jeremy Tinker et al. "Toward a Halo Mass Function for Precision Cosmology: The Limits of Universality". In: ApJ 688.2 (Dec. 2008), pp. 709–728. DOI: `10.1086/591439`. arXiv: `0803.2706 [astro-ph]`.

[60] Joseph DeRose et al. "The AEMULUS Project. I. Numerical Simulations for Precision Cosmology". In: ApJ 875.1, 69 (Apr. 2019), p. 69. DOI: `10.3847/1538-4357/ab1085`. arXiv: `1804.05865 [astro-ph.CO]`.

[61] Aurel Schneider et al. "Matter power spectrum and the challenge of percent accuracy". In: J. Cosmology Astropart. Phys. 2016.4, 047 (Apr. 2016), p. 047. DOI: `10.1088/1475-7516/2016/04/047`. arXiv: `1503.05920 [astro-ph.CO]`.

[62] Yu Feng 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]`.

[63] Svetlin Tassev, Matias Zaldarriaga, and Daniel J. Eisenstein. "Solving large scale structure in ten easy steps with COLA". In: J. Cosmology Astropart. Phys. 2013.6, 036 (June 2013), p. 036. DOI: `10.1088/1475-7516/2013/06/036`. arXiv: `1301.0322 [astro-ph.CO]`.

[64] Martin White, Jeremy L. Tinker, and Cameron K. McBride. "Mock galaxy catalogues using the quick particle mesh method". In: MNRAS 437.3 (Jan. 2014), pp. 2594–2606. DOI: `10.1093/mnras/stt2071`. arXiv: `1309.5532 [astro-ph.CO]`.

[65] 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]`.

[66] Earl Lawrence et al. "The Mira-Titan Universe. II. Matter Power Spectrum Emulation". In: ApJ 847.1, 50 (Sept. 2017), p. 50. DOI: `10.3847/1538-4357/aa86a9`. arXiv: `1705.03388 [astro-ph.CO]`.

[67] Takahiro Nishimichi et al. "Dark Quest. I. Fast and Accurate Emulation of Halo Clustering Statistics and Its Application to Galaxy Clustering". In: ApJ 884.1, 29 (Oct. 2019), p. 29. DOI: `10.3847/1538-4357/ab3719`. arXiv: `1811.09504 [astro-ph.CO]`.

[68] Ryuichi Takahashi et al. "Revising the Halofit Model for the Nonlinear Matter Power Spectrum". In: ApJ 761.2, 152 (Dec. 2012), p. 152. DOI: `10.1088/0004-637X/761/2/152`. arXiv: `1208.2701 [astro-ph.CO]`.

[69]   Martin White. "Baryons and weak lensing power spectra". In: *Astroparticle Physics* 22.2 (Nov. 2004), pp. 211–217. DOI: 10.1016/j.astropartphys.2004.06.001. arXiv: astro-ph/0405593 [astro-ph].

[70]   Hu Zhan and Lloyd Knox. "Effect of Hot Baryons on the Weak-Lensing Shear Power Spectrum". In: ApJ 616.2 (Dec. 2004), pp. L75–L78. DOI: 10.1086/426712. arXiv: astro-ph/0409198 [astro-ph].

[71]   Nora Elisa Chisari et al. "Modelling baryonic feedback for survey cosmology". In: *The Open Journal of Astrophysics* 2.1, 4 (June 2019), p. 4. DOI: 10.21105/astro.1905.06082. arXiv: 1905.06082 [astro-ph.CO].

[72]   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].

[73]   A. J. Mead et al. "An accurate halo model for fitting non-linear cosmological power spectra and baryonic feedback models". In: MNRAS 454.2 (Dec. 2015), pp. 1958–1975. DOI: 10.1093/mnras/stv2036. arXiv: 1505.07833 [astro-ph.CO].

[74]   Aurel Schneider and Romain Teyssier. "A new method to quantify the effects of baryons on the matter power spectrum". In: J. Cosmology Astropart. Phys. 2015.12, 049 (Dec. 2015), p. 049. DOI: 10.1088/1475-7516/2015/12/049. arXiv: 1510.06034 [astro-ph.CO].

[75]   Aurel Schneider et al. "Quantifying baryon effects on the matter power spectrum and the weak lensing shear correlation". In: J. Cosmology Astropart. Phys. 2019.3, 020 (Mar. 2019), p. 020. DOI: 10.1088/1475-7516/2019/03/020. arXiv: 1810.08629 [astro-ph.CO].

[76]   Zvonimir Vlah, Uroš Seljak, and Tobias Baldauf. "Lagrangian perturbation theory at one loop order: Successes, failures, and improvements". In: Phys. Rev. D 91.2, 023508 (Jan. 2015), p. 023508. DOI: 10.1103/PhysRevD.91.023508. arXiv: 1410.1617 [astro-ph.CO].

[77]   S. Pandey et al. "Perturbation theory for modeling galaxy bias: Validation with simulations of the Dark Energy Survey". In: Phys. Rev. D 102.12, 123522 (Dec. 2020), p. 123522. DOI: 10.1103/PhysRevD.102.123522. arXiv: 2008.05991 [astro-ph.CO].

[78]   Chirag Modi, Shi-Fan Chen, and Martin White. "Simulations and symmetries". In: MNRAS 492.4 (Mar. 2020), pp. 5754–5763. DOI: 10.1093/mnras/staa251. arXiv: 1910.07097 [astro-ph.CO].

[79]   Alexander Mead and Licia Verde. "Including beyond-linear halo bias in halo models". In: *arXiv e-prints*, arXiv:2011.08858 (Nov. 2020), arXiv:2011.08858. arXiv: 2011.08858 [astro-ph.CO].

[80] 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].

[81] Zhongxu Zhai et al. "The Clustering of Luminous Red Galaxies at z ∼ 0.7 from EBOSS and BOSS Data". In: ApJ 848.2, 76 (Oct. 2017), p. 76. DOI: 10.3847/1538-4357/aa8eee. arXiv: 1607.05383 [astro-ph.CO].

[82] Beth A. Reid et al. "A 2.5 per cent measurement of the growth rate from small-scale redshift space clustering of SDSS-III CMASS galaxies". In: MNRAS 444.1 (Oct. 2014), pp. 476–502. DOI: 10.1093/mnras/stu1391. arXiv: 1404.3742 [astro-ph.CO].

[83] Irshad Mohammed and Uroš Seljak. "Analytic model for the matter power spectrum, its covariance matrix and baryonic effects". In: MNRAS 445.4 (Dec. 2014), pp. 3382–3400. DOI: 10.1093/mnras/stu1972. arXiv: 1407.0060 [astro-ph.CO].

[84] P. Valageas and T. Nishimichi. "Combining perturbation theories with halo models". In: A&A 527, A87 (Mar. 2011), A87. DOI: 10.1051/0004-6361/201015685. arXiv: 1009.0597 [astro-ph.CO].

[85] Oliver H. E. Philcox, David N. Spergel, and Francisco Villaescusa-Navarro. "Effective halo model: Creating a physical and accurate model of the matter power spectrum and cluster counts". In: Phys. Rev. D 101.12, 123520 (June 2020), p. 123520. DOI: 10.1103/PhysRevD.101.123520. arXiv: 2004.09515 [astro-ph.CO].

[86] Patrick Valageas. "Accuracy of analytical models of the large-scale matter distribution". In: Phys. Rev. D 88.8, 083524 (Oct. 2013), p. 083524. DOI: 10.1103/PhysRevD.88.083524. arXiv: 1308.6755 [astro-ph.CO].

[87] M. Davis et al. "The evolution of large-scale structure in a universe dominated by cold dark matter". In: ApJ 292 (May 1985), pp. 371–394. DOI: 10.1086/163168.

[88] Nick Hand et al. "nbodykit: An Open-source, Massively Parallel Toolkit for Large-scale Structure". In: AJ 156.4, 160 (Oct. 2018), p. 160. DOI: 10.3847/1538-3881/aadae0. arXiv: 1712.05834 [astro-ph.IM].

[89] Y. P. Jing. "Correcting for the Alias Effect When Measuring the Power Spectrum Using a Fast Fourier Transform". In: ApJ 620.2 (Feb. 2005), pp. 559–563. DOI: 10.1086/427087. arXiv: astro-ph/0409240 [astro-ph].

[90] Manodeep Sinha and Lehman H. Garrison. "CORRFUNC - a suite of blazing fast correlation functions on the CPU". In: MNRAS 491.2 (Jan. 2020), pp. 3022–3041. DOI: 10.1093/mnras/stz3157. arXiv: 1911.03545 [astro-ph.CO].

[91] Francisco Villaescusa-Navarro et al. "Statistical Properties of Paired Fixed Fields". In: ApJ 867.2, 137 (Nov. 2018), p. 137. DOI: 10.3847/1538-4357/aae52b. arXiv: 1806.01871 [astro-ph.CO].

[92] Zvonimir Vlah, Emanuele Castorina, and Martin White. "The Gaussian streaming model and convolution Lagrangian effective field theory". In: J. Cosmology Astropart. Phys. 2016.12, 007 (Dec. 2016), p. 007. DOI: 10.1088/1475-7516/2016/12/007. arXiv: 1609.02908 [astro-ph.CO].

[93] Shi-Fan Chen, Zvonimir Vlah, and Martin White. "Consistent Modeling of Velocity Statistics and Redshift-Space Distortions in One-Loop Perturbation Theory". In: *arXiv e-prints*, arXiv:2005.00523 (May 2020), arXiv:2005.00523. arXiv: 2005.00523 [astro-ph.CO].

[94] Andrew P. Hearin et al. "Forward Modeling of Large-scale Structure: An Open-source Approach with Halotools". In: AJ 154.5, 190 (Nov. 2017), p. 190. DOI: 10.3847/1538-3881/aa859f. arXiv: 1606.04106 [astro-ph.IM].

[95] Zheng Zheng, Alison L. Coil, and Idit Zehavi. "Galaxy Evolution from Halo Occupation Distribution Modeling of DEEP2 and SDSS Galaxy Clustering". In: ApJ 667.2 (Oct. 2007), pp. 760–779. DOI: 10.1086/521074. arXiv: astro-ph/0703457 [astro-ph].

[96] Beth Reid et al. "SDSS-III Baryon Oscillation Spectroscopic Survey Data Release 12: galaxy target selection and large-scale structure catalogues". In: MNRAS 455.2 (Jan. 2016), pp. 1553–1573. DOI: 10.1093/mnras/stv2382. arXiv: 1509.06529 [astro-ph.CO].

[97] E. Sefusatti et al. "Accurate estimators of correlation functions in Fourier space". In: MNRAS 460.4 (Aug. 2016), pp. 3624–3636. DOI: 10.1093/mnras/stw1229. arXiv: 1512.07295 [astro-ph.CO].

[98] 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].

[99] Nick Hand et al. "Extending the modeling of the anisotropic galaxy power spectrum to k = 0.4 hMpc^{-1}". In: J. Cosmology Astropart. Phys. 2017.10, 009 (Oct. 2017), p. 009. DOI: 10.1088/1475-7516/2017/10/009. arXiv: 1706.02362 [astro-ph.CO].

[100] Julio F. Navarro, Carlos S. Frenk, and Simon D. M. White. "The Structure of Cold Dark Matter Halos". In: ApJ 462 (May 1996), p. 563. DOI: 10.1086/177173. arXiv: astro-ph/9508025 [astro-ph].

[101] Aaron A. Dutton and Andrea V. Macciò. "Cold dark matter haloes in the Planck era: evolution of structural parameters for Einasto and NFW profiles". In: MNRAS 441.4 (July 2014), pp. 3359–3374. DOI: 10.1093/mnras/stu742. arXiv: 1402.7073 [astro-ph.CO].

[102] Takahiko Matsubara. "Resumming cosmological perturbations via the Lagrangian picture: One-loop results in real space and in redshift space". In: Phys. Rev. D 77.6, 063530 (Mar. 2008), p. 063530. DOI: 10.1103/PhysRevD.77.063530. arXiv: 0711.2521 [astro-ph].

[103] Svetlin Tassev. "N-point statistics of large-scale structure in the Zel'dovich approximation". In: J. Cosmology Astropart. Phys. 2014.6, 012 (June 2014), p. 012. DOI: 10.1088/1475-7516/2014/06/012. arXiv: 1311.6316 [astro-ph.CO].

[104] P. J. E. Peebles. *The large-scale structure of the universe.* 1980.

[105] Alice Y. Chen and Niayesh Afshordi. "Amending the halo model to satisfy cosmological conservation laws". In: Phys. Rev. D 101.10, 103522 (May 2020), p. 103522. DOI: 10.1103/PhysRevD.101.103522. arXiv: 1912.04872 [astro-ph.CO].

[106] N. E. Chisari et al. "The impact of baryons on the matter power spectrum from the Horizon-AGN cosmological hydrodynamical simulation". In: MNRAS 480.3 (Nov. 2018), pp. 3962–3977. DOI: 10.1093/mnras/sty2093. arXiv: 1801.08559 [astro-ph.CO].

[107] Shi-Fan Chen, Emanuele Castorina, and Martin White. "Biased tracers of two fluids in the Lagrangian picture". In: J. Cosmology Astropart. Phys. 2019.6, 006 (June 2019), p. 006. DOI: 10.1088/1475-7516/2019/06/006. arXiv: 1903.00437 [astro-ph.CO].

[108] Fabian Schmidt. "Effect of relative velocity and density perturbations between baryons and dark matter on the clustering of galaxies". In: Phys. Rev. D 94.6, 063508 (Sept. 2016), p. 063508. DOI: 10.1103/PhysRevD.94.063508. arXiv: 1602.09059 [astro-ph.CO].

[109] Cornelius Rampf, Cora Uhlemann, and Oliver Hahn. "Cosmological perturbations for two cold fluids in ΛCDM". In: MNRAS 503.1 (May 2021), pp. 406–425. DOI: 10.1093/mnras/staa3605. arXiv: 2008.09123 [astro-ph.CO].

[110] Alexandre Barreira et al. "Baryon-CDM isocurvature galaxy bias with IllustrisTNG". In: J. Cosmology Astropart. Phys. 2020.2, 005 (Feb. 2020), p. 005. DOI: 10.1088/1475-7516/2020/02/005. arXiv: 1907.04317 [astro-ph.CO].

[111] M. A. Troxel 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].

[112] Marcel P. van Daalen et al. "The effects of galaxy formation on the matter power spectrum: a challenge for precision cosmology". In: MNRAS 415.4 (Aug. 2011), pp. 3649–3665. DOI: 10.1111/j.1365-2966.2011.18981.x. arXiv: 1104.1174 [astro-ph.CO].

[113] Mark Vogelsberger et al. "Introducing the Illustris Project: simulating the coevolution of dark and visible matter in the Universe". In: MNRAS 444.2 (Oct. 2014), pp. 1518–1547. DOI: 10.1093/mnras/stu1536. arXiv: 1405.2921 [astro-ph.CO].

[114] 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].

[115] Nishikanta Khandai et al. "The MassiveBlack-II simulation: the evolution of haloes and galaxies to z ∼ 0". In: MNRAS 450.2 (June 2015), pp. 1349–1374. DOI: `10.1093/mnras/stv627`. arXiv: `1402.0888 [astro-ph.CO]`.

[116] 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]`.

[117] 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]`.

[118] Hung-Jin Huang et al. "Modelling baryonic physics in future weak lensing surveys". In: MNRAS 488.2 (Sept. 2019), pp. 1652–1678. DOI: `10.1093/mnras/stz1714`. arXiv: `1809.01146 [astro-ph.CO]`.

[119] 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]`.

[120] Chirag Modi, Martin White, and Zvonimir Vlah. "Modeling CMB lensing cross correlations with CLEFT". In: J. Cosmology Astropart. Phys. 2017.8, 009 (Aug. 2017), p. 009. DOI: `10.1088/1475-7516/2017/08/009`. arXiv: `1706.03173 [astro-ph.CO]`.

[121] Jordan Carlson, Beth Reid, and Martin White. "Convolution Lagrangian perturbation theory for biased tracers". In: MNRAS 429.2 (Feb. 2013), pp. 1674–1685. DOI: `10.1093/mnras/sts457`. arXiv: `1209.0780 [astro-ph.CO]`.

[122] Svetlin Tassev and Matias Zaldarriaga. "Estimating CDM particle trajectories in the mildly non-linear regime of structure formation. Implications for the density field in real and redshift space". In: J. Cosmology Astropart. Phys. 2012.12, 011 (Dec. 2012), p. 011. DOI: `10.1088/1475-7516/2012/12/011`. arXiv: `1203.5785 [astro-ph.CO]`.

[123] Tobias Baldauf et al. "Halo stochasticity from exclusion and nonlinear clustering". In: Phys. Rev. D 88.8, 083507 (Oct. 2013), p. 083507. DOI: `10.1103/PhysRevD.88.083507`. arXiv: `1305.2917 [astro-ph.CO]`.

[124] Nico Hamaus et al. "Minimizing the stochasticity of halos in large-scale structure surveys". In: Phys. Rev. D 82.4, 043515 (Aug. 2010), p. 043515. DOI: `10.1103/PhysRevD.82.043515`. arXiv: `1004.5377 [astro-ph.CO]`.

[125] Marcel Schmittfull et al. "Modeling biased tracers at the field level". In: Phys. Rev. D 100.4, 043514 (Aug. 2019), p. 043514. DOI: `10.1103/PhysRevD.100.043514`. arXiv: `1811.10640 [astro-ph.CO]`.

[126] Robert E. Smith, Román Scoccimarro, and Ravi K. Sheth. "Scale dependence of halo and galaxy bias: Effects in real space". In: Phys. Rev. D 75.6, 063512 (Mar. 2007), p. 063512. DOI: `10.1103/PhysRevD.75.063512`. arXiv: `astro-ph/0609547 [astro-ph]`.

[127] Tobias Baldauf et al. "Peak exclusion, stochasticity and convergence of perturbative bias expansions in 1+1 gravity". In: MNRAS 456.4 (Mar. 2016), pp. 3985–4000. DOI: `10.1093/mnras/stv2973`. arXiv: `1510.09204 [astro-ph.CO]`.

[128] Tobias Baldauf et al. "Non-perturbative halo clustering from cosmological density peaks". In: *arXiv e-prints*, arXiv:2012.14404 (Dec. 2020), arXiv:2012.14404. arXiv: `2012.14404 [astro-ph.CO]`.

[129] Frank C. van den Bosch et al. "Cosmological constraints from a combination of galaxy clustering and lensing - I. Theoretical framework". In: MNRAS 430.2 (Apr. 2013), pp. 725–746. DOI: `10.1093/mnras/sts006`. arXiv: `1206.6890 [astro-ph.CO]`.

[130] Zvonimir Vlah et al. "Distribution function approach to redshift space distortions. Part V: perturbation theory applied to dark matter halos". In: J. Cosmology Astropart. Phys. 2013.10, 053 (Oct. 2013), p. 053. DOI: `10.1088/1475-7516/2013/10/053`. arXiv: `1308.6294 [astro-ph.CO]`.

[131] Teppei Okumura et al. "Galaxy power spectrum in redshift space: Combining perturbation theory with the halo model". In: Phys. Rev. D 92.10, 103516 (Nov. 2015), p. 103516. DOI: `10.1103/PhysRevD.92.103516`. arXiv: `1506.05814 [astro-ph.CO]`.

[132] Chiaki Hikage, Masahiro Takada, and David N. Spergel. "Using galaxy-galaxy weak lensing measurements to correct the finger of God". In: MNRAS 419.4 (Feb. 2012), pp. 3457–3481. DOI: `10.1111/j.1365-2966.2011.19987.x`. arXiv: `1106.1640 [astro-ph.CO]`.

[133] Surhud More et al. "The Weak Lensing Signal and the Clustering of BOSS Galaxies. II. Astrophysical and Cosmological Constraints". In: ApJ 806.1, 2 (June 2015), p. 2. DOI: `10.1088/0004-637X/806/1/2`. arXiv: `1407.1856 [astro-ph.CO]`.

[134] Martin White. "The Mass Function". In: ApJS 143.2 (Dec. 2002), pp. 241–255. DOI: `10.1086/342752`. arXiv: `astro-ph/0207185 [astro-ph]`.

[135] Rachel Mandelbaum et al. "Galaxy halo masses and satellite fractions from galaxy-galaxy lensing in the Sloan Digital Sky Survey: stellar mass, luminosity, morphology and environment dependencies". In: MNRAS 368.2 (May 2006), pp. 715–731. DOI: `10.1111/j.1365-2966.2006.10156.x`. arXiv: `astro-ph/0511164 [astro-ph]`.

[136] Benjamin D. Wibking et al. "Cosmology with galaxy-galaxy lensing on non-perturbative scales: emulation method and application to BOSS LOWZ". In: MNRAS 492.2 (Feb. 2020), pp. 2872–2896. DOI: `10.1093/mnras/stz3423`. arXiv: `1907.06293 [astro-ph.CO]`.

[137] John K. Parejko et al. "The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: the low-redshift sample". In: MNRAS 429.1 (Feb. 2013), pp. 98–112. DOI: `10.1093/mnras/sts314`. arXiv: `1211.3976 [astro-ph.CO]`.

[138] Biwei Dai and Uros Seljak. "Sliced Iterative Generator". In: *arXiv e-prints*, arXiv:2007.00674 (July 2020), arXiv:2007.00674. arXiv: `2007.00674 [cs.LG]`.

[139] Juliana Kwan et al. "Cosmic Emulation: Fast Predictions for the Galaxy Power Spectrum". In: ApJ 810.1, 35 (Sept. 2015), p. 35. DOI: 10.1088/0004-637X/810/1/35. arXiv: 1311.6444 [astro-ph.CO].

[140] Nickolas Kokron et al. "The cosmology dependence of galaxy clustering and lensing from a hybrid N-body-perturbation theory model". In: *arXiv e-prints*, arXiv:2101.11014 (Jan. 2021), arXiv:2101.11014. arXiv: 2101.11014 [astro-ph.CO].

[141] Hironao Miyatake et al. "Cosmological inference from emulator based halo model I: Validation tests with HSC and SDSS mock catalogs". In: *arXiv e-prints*, arXiv:2101.00113 (Dec. 2020), arXiv:2101.00113. arXiv: 2101.00113 [astro-ph.CO].

[142] Keir K. Rogers et al. "Bayesian emulator optimisation for cosmology: application to the Lyman-alpha forest". In: J. Cosmology Astropart. Phys. 2019.2, 031 (Feb. 2019), p. 031. DOI: 10.1088/1475-7516/2019/02/031. arXiv: 1812.04631 [astro-ph.CO].

[143] Yosuke Kobayashi et al. "Accurate emulator for the redshift-space power spectrum of dark matter halos and its application to galaxy power spectrum". In: Phys. Rev. D 102.6, 063504 (Sept. 2020), p. 063504. DOI: 10.1103/PhysRevD.102.063504. arXiv: 2005.06122 [astro-ph.CO].

[144] David Valcin et al. "BE-HaPPY: bias emulator for halo power spectrum including massive neutrinos". In: J. Cosmology Astropart. Phys. 2019.12, 057 (Dec. 2019), p. 057. DOI: 10.1088/1475-7516/2019/12/057. arXiv: 1901.06045 [astro-ph.CO].

[145] Ying Zu. "On the "Lensing is Low" of BOSS Galaxies". In: *arXiv e-prints*, arXiv:2010.01143 (Oct. 2020), arXiv:2010.01143. arXiv: 2010.01143 [astro-ph.CO].

[146] Stéfan van der Walt, S. Chris Colbert, and Gaël Varoquaux. "The NumPy Array: A Structure for Efficient Numerical Computation". In: *Computing in Science and Engineering* 13.2 (Mar. 2011), pp. 22–30. DOI: 10.1109/MCSE.2011.37. arXiv: 1102.1523 [cs.MS].

[147] Pauli Virtanen et al. "SciPy 1.0: fundamental algorithms for scientific computing in Python". In: *Nature Methods* 17 (Feb. 2020), pp. 261–272. DOI: 10.1038/s41592-019-0686-2. arXiv: 1907.10121 [cs.MS].

[148] Astropy Collaboration et al. "The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package". In: AJ 156.3, 123 (Sept. 2018), p. 123. DOI: 10.3847/1538-3881/aabc4f. arXiv: 1801.02634 [astro-ph.IM].

[149] Matteo Frigo and Steven G. Johnson. *FFTW: Fastest Fourier Transform in the West*. Jan. 2012. ascl: 1201.015.

[150] Antony Lewis, Anthony Challinor, and Anthony Lasenby. "Efficient Computation of Cosmic Microwave Background Anisotropies in Closed Friedmann-Robertson-Walker Models". In: ApJ 538.2 (Aug. 2000), pp. 473–476. DOI: 10.1086/309179. arXiv: astro-ph/9911177 [astro-ph].

[151] Yin Li. *mcfit: Multiplicatively Convolutional Fast Integral Transforms*. June 2019. ascl: `1906.017`.

[152] Aurel Schneider et al. "Baryonic effects for weak lensing. Part I. Power spectrum and covariance matrix". In: J. Cosmology Astropart. Phys. 2020.4, 019 (Apr. 2020), p. 019. DOI: `10.1088/1475-7516/2020/04/019`. arXiv: `1910.11357 [astro-ph.CO]`.

[153] Joseph E. McEwen et al. "FAST-PT: a novel algorithm to calculate convolution integrals in cosmological perturbation theory". In: J. Cosmology Astropart. Phys. 2016.9, 015 (Sept. 2016), p. 015. DOI: `10.1088/1475-7516/2016/09/015`. arXiv: `1603.04826 [astro-ph.CO]`.

[154] M. White. "The mass of a halo". In: A&A 367 (Feb. 2001), pp. 27–32. DOI: `10.1051/0004-6361:20000357`. arXiv: `astro-ph/0011495 [astro-ph]`.

[155] Zarija Lukić et al. "The Structure of Halos: Implications for Group and Cluster Cosmology". In: ApJ 692.1 (Feb. 2009), pp. 217–228. DOI: `10.1088/0004-637X/692/1/217`. arXiv: `0803.3624 [astro-ph]`.

[156] Rafael García and Eduardo Rozo. "Halo exclusion criteria impacts halo statistics". In: MNRAS 489.3 (Nov. 2019), pp. 4170–4175. DOI: `10.1093/mnras/stz2458`. arXiv: `1903.01709 [astro-ph.CO]`.

[157] A Baha Balantekin and Boris Kayser. "On the properties of neutrinos". In: *Annual Review of Nuclear and Particle Science* 68 (2018), pp. 313–338.

[158] André De Gouvêa. "Neutrino mass models". In: *Annual Review of Nuclear and Particle Science* 66 (2016), pp. 197–217.

[159] Cora Dvorkin et al. "Neutrino Mass from Cosmology: Probing Physics Beyond the Standard Model". In: BAAS 51.3, 64 (May 2019), p. 64. arXiv: `1903.03689 [astro-ph.CO]`.

[160] Katrin Collaboration et al. "Direct neutrino-mass measurement with sub-electronvolt sensitivity". In: *Nature Physics* 18.2 (Feb. 2022), pp. 160–166. DOI: `10.1038/s41567-021-01463-1`.

[161] Shadab Alam et al. "Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Cosmological implications from two decades of spectroscopic surveys at the Apache Point Observatory". In: Phys. Rev. D 103.8, 083533 (Apr. 2021), p. 083533. DOI: `10.1103/PhysRevD.103.083533`. arXiv: `2007.08991 [astro-ph.CO]`.

[162] P. F. de Salas et al. "Status of neutrino oscillations 2018: first hint for normal mass ordering and improved CP sensitivity". In: *arXiv e-prints*, arXiv:1708.01186 (Aug. 2017), arXiv:1708.01186. arXiv: `1708.01186 [hep-ph]`.

[163] Ivan Esteban et al. "Global analysis of three-flavour neutrino oscillations: synergies and tensions in the determination of θ_{23}, δ_{CP}, and the mass ordering". In: *Journal of High Energy Physics* 2019.1, 106 (Jan. 2019), p. 106. DOI: `10.1007/JHEP01(2019)106`. arXiv: `1811.05487 [hep-ph]`.

[164] R. G. H. Robertson et al. "Limit on nu-bar $_e$ mass from observation of the β decay of molecular tritium". In: Phys. Rev. Lett. 67.8 (Aug. 1991), pp. 957–960. DOI: 10.1103/PhysRevLett.67.957.

[165] V. M. Lobashev et al. "Neutrino mass and anomaly in the tritium beta-spectrum. Results of the "Troitsk ν-mass" experiment". In: *Nuclear Physics B Proceedings Supplements* 77.1 (May 1999), pp. 327–332. DOI: 10.1016/S0920-5632(99)00438-7.

[166] Ch Kraus et al. "Final results from phase II of the Mainz neutrino mass searchin tritium $\{\beta\}$ decay". In: *European Physical Journal C* 40.4 (Apr. 2005), pp. 447–468. DOI: 10.1140/epjc/s2005-02139-7. arXiv: hep-ex/0412056 [hep-ex].

[167] M. Aker et al. "KATRIN: status and prospects for the neutrino mass and beyond". In: *Journal of Physics G Nuclear Physics* 49.10, 100501 (Oct. 2022), p. 100501. DOI: 10.1088/1361-6471/ac834e. arXiv: 2203.08059 [nucl-ex].

[168] Maria Archidiacono, Steen Hannestad, and Julien Lesgourgues. "What will it take to measure individual neutrino mass states using cosmology?" In: J. Cosmology Astropart. Phys. 2020.9, 021 (Sept. 2020), p. 021. DOI: 10.1088/1475-7516/2020/09/021. arXiv: 2003.03354 [astro-ph.CO].

[169] 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. arXiv: 1611.00036 [astro-ph.IM].

[170] Weishuang Linda Xu et al. "Accurately weighing neutrinos with cosmological surveys". In: Phys. Rev. D 103.2, 023503 (Jan. 2021), p. 023503. DOI: 10.1103/PhysRevD.103.023503. arXiv: 2006.09395 [astro-ph.CO].

[171] Florian Beutler et al. "The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: signs of neutrino mass in current cosmological data sets". In: MNRAS 444.4 (Nov. 2014), pp. 3501–3516. DOI: 10.1093/mnras/stu1702. arXiv: 1403.4599 [astro-ph.CO].

[172] Nathalie Palanque-Delabrouille et al. "Neutrino masses and cosmology with Lyman-alpha forest power spectrum". In: J. Cosmology Astropart. Phys. 2015.11 (Nov. 2015), pp. 011–011. DOI: 10.1088/1475-7516/2015/11/011. arXiv: 1506.05976 [astro-ph.CO].

[173] Antonio J. Cuesta, Viviana Niro, and Licia Verde. "Neutrino mass limits: Robust information from the power spectrum of galaxy surveys". In: *Physics of the Dark Universe* 13 (Sept. 2016), pp. 77–86. DOI: 10.1016/j.dark.2016.04.005. arXiv: 1511.05983 [astro-ph.CO].

[174] Sunny Vagnozzi et al. "Unveiling ν secrets with cosmological data: Neutrino masses and mass hierarchy". In: Phys. Rev. D 96.12, 123503 (Dec. 2017), p. 123503. DOI: 10.1103/PhysRevD.96.123503. arXiv: 1701.08172 [astro-ph.CO].

[175] Cyrille Doux et al. "Cosmological constraints from a joint analysis of cosmic microwave background and spectroscopic tracers of the large-scale structure". In: MNRAS 480.4 (Nov. 2018), pp. 5386–5411. DOI: 10.1093/mnras/sty2160. arXiv: 1706.04583 [astro-ph.CO].

[176] Amol Upadhye. "Neutrino mass and dark energy constraints from redshift-space distortions". In: J. Cosmology Astropart. Phys. 2019.5, 041 (May 2019), p. 041. DOI: 10.1088/1475-7516/2019/05/041. arXiv: 1707.09354 [astro-ph.CO].

[177] Mikhail M. Ivanov, Marko Simonović, and Matias Zaldarriaga. "Cosmological parameters and neutrino masses from the final P l a n c k and full-shape BOSS data". In: Phys. Rev. D 101.8, 083504 (Apr. 2020), p. 083504. DOI: 10.1103/PhysRevD.101.083504. arXiv: 1912.08208 [astro-ph.CO].

[178] Nathalie Palanque-Delabrouille et al. "Hints, neutrino bounds, and WDM constraints from SDSS DR14 Lyman-α and Planck full-survey data". In: J. Cosmology Astropart. Phys. 2020.4, 038 (Apr. 2020), p. 038. DOI: 10.1088/1475-7516/2020/04/038. arXiv: 1911.09073 [astro-ph.CO].

[179] 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].

[180] Arka Banerjee et al. "Weighing neutrinos with the halo environment". In: J. Cosmology Astropart. Phys. 2020.6, 032 (June 2020), p. 032. DOI: 10.1088/1475-7516/2020/06/032. arXiv: 1907.06598 [astro-ph.CO].

[181] Sunny Vagnozzi et al. "Bias due to neutrinos must not uncorrect'd go". In: J. Cosmology Astropart. Phys. 2018.9, 001 (Sept. 2018), p. 001. DOI: 10.1088/1475-7516/2018/09/001. arXiv: 1807.04672 [astro-ph.CO].

[182] Nico Schuster et al. "The bias of cosmic voids in the presence of massive neutrinos". In: J. Cosmology Astropart. Phys. 2019.12, 055 (Dec. 2019), p. 055. DOI: 10.1088/1475-7516/2019/12/055. arXiv: 1905.00436 [astro-ph.CO].

[183] Christina D. Kreisch et al. "Massive neutrinos leave fingerprints on cosmic voids". In: MNRAS 488.3 (Sept. 2019), pp. 4413–4426. DOI: 10.1093/mnras/stz1944. arXiv: 1808.07464 [astro-ph.CO].

[184] Elena Massara et al. "Using the Marked Power Spectrum to Detect the Signature of Neutrinos in Large-Scale Structure". In: Phys. Rev. Lett. 126.1, 011301 (Jan. 2021), p. 011301. DOI: 10.1103/PhysRevLett.126.011301. arXiv: 2001.11024 [astro-ph.CO].

[185] Adrian E. Bayer, Arka Banerjee, and Uroš Seljak. "Beware of fake ν 's: The effect of massive neutrinos on the nonlinear evolution of cosmic structure". In: Phys. Rev. D 105.12, 123510 (June 2022), p. 123510. DOI: 10.1103/PhysRevD.105.123510. arXiv: 2108.04215 [astro-ph.CO].

[186] ChangHoon Hahn et al. "Constraining M_ν with the bispectrum. Part I. Breaking parameter degeneracies". In: J. Cosmology Astropart. Phys. 2020.3, 040 (Mar. 2020), p. 040. DOI: 10.1088/1475-7516/2020/03/040. arXiv: 1909.11107 [astro-ph.CO].

[187] Rossana Ruggeri et al. "DEMNUni: massive neutrinos and the bispectrum of large scale structures". In: J. Cosmology Astropart. Phys. 2018.3, 003 (Mar. 2018), p. 003. DOI: 10.1088/1475-7516/2018/03/003. arXiv: 1712.02334 [astro-ph.CO].

[188] William R. Coulton 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].

[189] Kevork Abazajian et al. "Nonlinear cosmological matter power spectrum with massive neutrinos: The halo model". In: Phys. Rev. D 71.4, 043507 (Feb. 2005), p. 043507. DOI: 10.1103/PhysRevD.71.043507. arXiv: astro-ph/0411552 [astro-ph].

[190] Steen Hannestad, Amol Upadhye, and Yvonne Y. Y. Wong. "Spoon or slide? The nonlinear matter power spectrum in the presence of massive neutrinos". In: J. Cosmology Astropart. Phys. 2020.11, 062 (Nov. 2020), p. 062. DOI: 10.1088/1475-7516/2020/11/062. arXiv: 2006.04995 [astro-ph.CO].

[191] Shankar Agarwal and Hume A. Feldman. "The effect of massive neutrinos on the matter power spectrum". In: MNRAS 410.3 (Jan. 2011), pp. 1647–1654. DOI: 10.1111/j.1365-2966.2010.17546.x. arXiv: 1006.0689 [astro-ph.CO].

[192] Amol Upadhye et al. "Large-scale structure formation with massive neutrinos and dynamical dark energy". In: Phys. Rev. D 89.10, 103515 (May 2014), p. 103515. DOI: 10.1103/PhysRevD.89.103515. arXiv: 1309.5872 [astro-ph.CO].

[193] Nina A. Maksimova et al. "ABACUSSUMMIT: a massive set of high-accuracy, high-resolution N-body simulations". In: MNRAS 508.3 (Dec. 2021), pp. 4017–4037. DOI: 10.1093/mnras/stab2484. arXiv: 2110.11398 [astro-ph.CO].

[194] Kelly R. Moran et al. "The Mira-Titan Universe IV. High Precision Power Spectrum Emulation". In: *arXiv e-prints*, arXiv:2207.12345 (July 2022), arXiv:2207.12345. arXiv: 2207.12345 [astro-ph.CO].

[195] Jacob Brandbyge and Steen Hannestad. "Grid based linear neutrino perturbations in cosmological N-body simulations". In: J. Cosmology Astropart. Phys. 2009.5, 002 (May 2009), p. 002. DOI: 10.1088/1475-7516/2009/05/002. arXiv: 0812.3149 [astro-ph].

[196] Diego Blas et al. "Structure formation with massive neutrinos: going beyond linear theory". In: J. Cosmology Astropart. Phys. 2014.11, 039 (Nov. 2014), p. 039. DOI: 10.1088/1475-7516/2014/11/039. arXiv: 1408.2995 [astro-ph.CO].

[197] Yacine Ali-Haïmoud and Simeon Bird. "An efficient implementation of massive neutrinos in non-linear structure formation simulations". In: MNRAS 428.4 (Feb. 2013), pp. 3375–3389. DOI: 10.1093/mnras/sts286. arXiv: 1209.0461 [astro-ph.CO].

[198] Jia Liu 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]`.

[199] Joe Zhiyu Chen, Amol Upadhye, and Yvonne Y. Y. Wong. "One line to run them all: SuperEasy massive neutrino linear response in N-body simulations". In: *arXiv e-prints*, arXiv:2011.12504 (Nov. 2020), arXiv:2011.12504. arXiv: `2011.12504 [astro-ph.CO]`.

[200] Kohji Yoshikawa et al. "Cosmological VlasovPoisson Simulations of Structure Formation with Relic Neutrinos: Nonlinear Clustering and the Neutrino Mass". In: ApJ 904.2, 159 (Dec. 2020), p. 159. DOI: `10.3847/1538-4357/abbd46`. arXiv: `2010.00248 [astro-ph.CO]`.

[201] Arka Banerjee and Neal Dalal. "Simulating nonlinear cosmological structure formation with massive neutrinos". In: J. Cosmology Astropart. Phys. 2016.11, 015 (Nov. 2016), p. 015. DOI: `10.1088/1475-7516/2016/11/015`. arXiv: `1606.06167 [astro-ph.CO]`.

[202] Jeppe Dakin et al. "νCONCEPT: cosmological neutrino simulations from the nonlinear Boltzmann hierarchy". In: J. Cosmology Astropart. Phys. 2019.2, 052 (Feb. 2019), p. 052. DOI: `10.1088/1475-7516/2019/02/052`. arXiv: `1712.03944 [astro-ph.CO]`.

[203] Derek Inman and Hao-Ran Yu. "Simulating the Cosmic Neutrino Background Using Collisionless Hydrodynamics". In: ApJS 250.1, 21 (Sept. 2020), p. 21. DOI: `10.3847/1538-4365/aba0b3`. arXiv: `2002.04601 [astro-ph.CO]`.

[204] Anatoly Klypin et al. "Structure Formation with Cold plus Hot Dark Matter". In: ApJ 416 (Oct. 1993), p. 1. DOI: `10.1086/173210`. arXiv: `astro-ph/9305011 [astro-ph]`.

[205] A Gardini, SA Bonometto, and G Murante. "Cluster mass function in mixed models". In: *The Astrophysical Journal* 524.2 (1999), p. 510.

[206] Jacob Brandbyge et al. "The effect of thermal neutrino motion on the non-linear cosmological matter power spectrum". In: J. Cosmology Astropart. Phys. 2008.8, 020 (Aug. 2008), p. 020. DOI: `10.1088/1475-7516/2008/08/020`. arXiv: `0802.3700 [astro-ph]`.

[207] Matteo Viel, Martin G. Haehnelt, and Volker Springel. "The effect of neutrinos on the matter distribution as probed by the intergalactic medium". In: J. Cosmology Astropart. Phys. 2010.6, 015 (June 2010), p. 015. DOI: `10.1088/1475-7516/2010/06/015`. arXiv: `1003.2422 [astro-ph.CO]`.

[208] Simeon Bird, Matteo Viel, and Martin G. Haehnelt. "Massive neutrinos and the nonlinear matter power spectrum". In: MNRAS 420.3 (Mar. 2012), pp. 2551–2561. DOI: `10.1111/j.1365-2966.2011.20222.x`. arXiv: `1109.4416 [astro-ph.CO]`.

[209] Francisco Villaescusa-Navarro et al. "Cosmology with massive neutrinos I: towards a realistic modeling of the relation between matter, haloes and galaxies". In: J. Cosmology Astropart. Phys. 2014.3, 011 (Mar. 2014), p. 011. DOI: 10.1088/1475-7516/2014/03/011. arXiv: 1311.0866 [astro-ph.CO].

[210] Emanuele Castorina et al. "DEMNUni: the clustering of large-scale structures in the presence of massive neutrinos". In: J. Cosmology Astropart. Phys. 2015.7, 043 (July 2015), p. 043. DOI: 10.1088/1475-7516/2015/07/043. arXiv: 1505.07148 [astro-ph.CO].

[211] Derek Inman et al. "Precision reconstruction of the cold dark matter-neutrino relative velocity from N -body simulations". In: Phys. Rev. D 92.2, 023502 (July 2015), p. 023502. DOI: 10.1103/PhysRevD.92.023502. arXiv: 1503.07480 [astro-ph.CO].

[212] J. D. Emberson et al. "Cosmological neutrino simulations at extreme scale". In: *Research in Astronomy and Astrophysics* 17.8, 085 (Aug. 2017), p. 085. DOI: 10.1088/1674-4527/17/8/85. arXiv: 1611.01545 [astro-ph.CO].

[213] Jacob Brandbyge, Steen Hannestad, and Thomas Tram. "Momentum space sampling of neutrinos in N-body simulations". In: J. Cosmology Astropart. Phys. 2019.3, 047 (Mar. 2019), p. 047. DOI: 10.1088/1475-7516/2019/03/047. arXiv: 1806.05874 [astro-ph.CO].

[214] Francisco Villaescusa-Navarro et al. "The Quijote Simulations". In: ApJS 250.1, 2 (Sept. 2020), p. 2. DOI: 10.3847/1538-4365/ab9d82. arXiv: 1909.05273 [astro-ph.CO].

[215] Adrian E. Bayer, Arka Banerjee, and Yu Feng. "A fast particle-mesh simulation of non-linear cosmological structure formation with massive neutrinos". In: J. Cosmology Astropart. Phys. 2021.1, 016 (Jan. 2021), p. 016. DOI: 10.1088/1475-7516/2021/01/016. arXiv: 2007.13394 [astro-ph.CO].

[216] Willem Elbers et al. "An optimal non-linear method for simulating relic neutrinos". In: MNRAS 507.2 (Oct. 2021), pp. 2614–2631. DOI: 10.1093/mnras/stab2260. arXiv: 2010.07321 [astro-ph.CO].

[217] J. Adamek et al. "Euclid: Modelling massive neutrinos in cosmology – a code comparison". In: *arXiv e-prints*, arXiv:2211.12457 (Nov. 2022), arXiv:2211.12457. arXiv: 2211.12457 [astro-ph.CO].

[218] Jacob Brandbyge and Steen Hannestad. "Resolving cosmic neutrino structure: a hybrid neutrino N-body scheme". In: J. Cosmology Astropart. Phys. 2010.1, 021 (Jan. 2010), p. 021. DOI: 10.1088/1475-7516/2010/01/021. arXiv: 0908.1969 [astro-ph.CO].

[219] Simeon Bird et al. "An efficient and accurate hybrid method for simulating non-linear neutrino structure". In: MNRAS 481.2 (Dec. 2018), pp. 1486–1500. DOI: 10.1093/mnras/sty2376. arXiv: 1803.09854 [astro-ph.CO].

[220] Joe Zhiyu Chen et al. "Hybrid multi-fluid-particle simulations of the cosmic neutrino background". In: *arXiv e-prints*, arXiv:2210.16012 (Oct. 2022), arXiv:2210.16012. arXiv: 2210.16012 [astro-ph.CO].

[221] Francisco Villaescusa-Navarro et al. "The Imprint of Neutrinos on Clustering in Redshift Space". In: ApJ 861.1, 53 (July 2018), p. 53. DOI: 10.3847/1538-4357/aac6bf. arXiv: 1708.01154 [astro-ph.CO].

[222] Benjamin O. Mummery et al. "The separate and combined effects of baryon physics and neutrino free streaming on large-scale structure". In: MNRAS 471.1 (Oct. 2017), pp. 227–242. DOI: 10.1093/mnras/stx1469. arXiv: 1702.02064 [astro-ph.CO].

[223] Graziano Rossi. "The Sejong Suite: Cosmological Hydrodynamical Simulations with Massive Neutrinos, Dark Radiation, and Warm Dark Matter". In: ApJS 249.2, 19 (Aug. 2020), p. 19. DOI: 10.3847/1538-4365/ab9d1e. arXiv: 2007.15279 [astro-ph.CO].

[224] Arka Banerjee et al. "Reducing noise in cosmological N-body simulations with neutrinos". In: J. Cosmology Astropart. Phys. 2018.9, 028 (Sept. 2018), p. 028. DOI: 10.1088/1475-7516/2018/09/028. arXiv: 1801.03906 [astro-ph.CO].

[225] Thomas Tram et al. "Fully relativistic treatment of light neutrinos in N-body simulations". In: J. Cosmology Astropart. Phys. 2019.3, 022 (Mar. 2019), p. 022. DOI: 10.1088/1475-7516/2019/03/022. arXiv: 1811.00904 [astro-ph.CO].

[226] Julian Adamek, Ruth Durrer, and Martin Kunz. "Relativistic N-body simulations with massive neutrinos". In: J. Cosmology Astropart. Phys. 2017.11, 004 (Nov. 2017), p. 004. DOI: 10.1088/1475-7516/2017/11/004. arXiv: 1707.06938 [astro-ph.CO].

[227] Christian Fidler et al. "Relativistic initial conditions for N-body simulations". In: J. Cosmology Astropart. Phys. 2017.6, 043 (June 2017), p. 043. DOI: 10.1088/1475-7516/2017/06/043. arXiv: 1702.03221 [astro-ph.CO].

[228] M. Zennaro et al. "Initial conditions for accurate N-body simulations of massive neutrino cosmologies". In: MNRAS 466.3 (Apr. 2017), pp. 3244–3258. DOI: 10.1093/mnras/stw3340. arXiv: 1605.05283 [astro-ph.CO].

[229] Salman Habib et al. "HACC: Simulating sky surveys on state-of-the-art supercomputing architectures". In: New A 42 (Jan. 2016), pp. 49–65. DOI: 10.1016/j.newast.2015.06.003. arXiv: 1410.2805 [astro-ph.IM].

[230] Antony Lewis, Anthony Challinor, and Anthony Lasenby. "Efficient Computation of Cosmic Microwave Background Anisotropies in Closed Friedmann-Robertson-Walker Models". In: ApJ 538.2 (Aug. 2000), pp. 473–476. DOI: 10.1086/309179. arXiv: astro-ph/9911177 [astro-ph].

[231] 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].

[232] P. J. E. Peebles. *Principles of Physical Cosmology.* 2021. DOI: `10.1515/9780691206721`.

[233] Adrian E. Bayer et al. "Detecting Neutrino Mass by Combining Matter Clustering, Halos, and Voids". In: ApJ 919.1, 24 (Sept. 2021), p. 24. DOI: `10.3847/1538-4357/ac0e91`. arXiv: `2102.05049 [astro-ph.CO]`.

[234] Carlo Giocoli, Marco Baldi, and Lauro Moscardini. "Weak lensing light-cones in modified gravity simulations with and without massive neutrinos". In: MNRAS 481.2 (Dec. 2018), pp. 2813–2828. DOI: `10.1093/mnras/sty2465`. arXiv: `1806.04681 [astro-ph.CO]`.

[235] Alvise Raccanelli, Licia Verde, and Francisco Villaescusa-Navarro. "Biases from neutrino bias: to worry or not to worry?" In: MNRAS 483.1 (Feb. 2019), pp. 734–743. DOI: `10.1093/mnras/sty2162`. arXiv: `1704.07837 [astro-ph.CO]`.

[236] Christian Fidler and Alexander Kleinjohann. "Suitable initial conditions for Newtonian simulations with massive neutrinos". In: J. Cosmology Astropart. Phys. 2019.6, 018 (June 2019), p. 018. DOI: `10.1088/1475-7516/2019/06/018`. arXiv: `1810.12019 [astro-ph.CO]`.

[237] G. Parimbelli et al. "The effects of massive neutrinos on the linear point of the correlation function". In: J. Cosmology Astropart. Phys. 2021.1, 009 (Jan. 2021), p. 009. DOI: `10.1088/1475-7516/2021/01/009`. arXiv: `2007.10345 [astro-ph.CO]`.

[238] S. Contreras et al. "3 per cent-accurate predictions for the clustering of dark matter, haloes, and subhaloes, over a wide range of cosmologies and scales". In: MNRAS 499.4 (Dec. 2020), pp. 4905–4917. DOI: `10.1093/mnras/staa3117`. arXiv: `2001.03176 [astro-ph.CO]`.

[239] Nicholas Frontiere et al. "Simulating Hydrodynamics in Cosmology with CRK-HACC". In: ApJS 264.2, 34 (Feb. 2023), p. 34. DOI: `10.3847/1538-4365/aca58d`. arXiv: `2202.02840 [astro-ph.CO]`.

[240] K. M. Górski et al. "HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere". In: ApJ 622.2 (Apr. 2005), pp. 759–771. DOI: `10.1086/427976`. arXiv: `astro-ph/0409513 [astro-ph]`.

[241] Naoki Yoshida, Naoshi Sugiyama, and Lars Hernquist. "The evolution of baryon density fluctuations in multicomponent cosmological simulations". In: MNRAS 344.2 (Sept. 2003), pp. 481–491. DOI: `10.1046/j.1365-8711.2003.06829.x`. arXiv: `astro-ph/0305210 [astro-ph]`.

[242] Caio Bastos de Senna Nascimento and Marilena Loverde. "Neutrinos in N-body simulations". In: Phys. Rev. D 104.4, 043512 (Aug. 2021), p. 043512. DOI: `10.1103/PhysRevD.104.043512`. arXiv: `2102.05690 [astro-ph.CO]`.

[243] Alejandro Aviles and Arka Banerjee. "A Lagrangian perturbation theory in the presence of massive neutrinos". In: J. Cosmology Astropart. Phys. 2020.10, 034 (Oct. 2020), p. 034. DOI: `10.1088/1475-7516/2020/10/034`. arXiv: `2007.06508` `[astro-ph.CO]`.

[244] 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]`.

[245] Willem Elbers et al. "Higher order initial conditions with massive neutrinos". In: MNRAS 516.3 (Nov. 2022), pp. 3821–3836. DOI: `10.1093/mnras/stac2365`. arXiv: `2202.00670` `[astro-ph.CO]`.

[246] M. Joyce and B. Marcos. "Quantification of discreteness effects in cosmological N-body simulations: Initial conditions". In: Phys. Rev. D 75.6, 063516 (Mar. 2007), p. 063516. DOI: `10.1103/PhysRevD.75.063516`. arXiv: `astro-ph/0410451` `[astro-ph]`.

[247] M. Joyce, B. Marcos, and T. Baertschiger. "Towards quantitative control on discreteness error in the non-linear regime of cosmological N-body simulations". In: MNRAS 394.2 (Apr. 2009), pp. 751–773. DOI: `10.1111/j.1365-2966.2008.14290.x`. arXiv: `0805.1357` `[astro-ph]`.

[248] Katrin Heitmann et al. "The Mira-Titan Universe: Precision Predictions for Dark Energy Surveys". In: ApJ 820.2, 108 (Apr. 2016), p. 108. DOI: `10.3847/0004-637X/820/2/108`. arXiv: `1508.02654` `[astro-ph.CO]`.

[249] 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]`.

[250] Roland de Putter et al. "Thinking outside the box: effects of modes larger than the survey on matter power spectrum covariance". In: J. Cosmology Astropart. Phys. 2012.4, 019 (Apr. 2012), p. 019. DOI: `10.1088/1475-7516/2012/04/019`. arXiv: `1111.6596` `[astro-ph.CO]`.

[251] Masahiro Takada and Wayne Hu. "Power spectrum super-sample covariance". In: Phys. Rev. D 87.12, 123504 (June 2013), p. 123504. DOI: `10.1103/PhysRevD.87.123504`. arXiv: `1302.6994` `[astro-ph.CO]`.

[252] Yin Li, Wayne Hu, and Masahiro Takada. "Super-sample signal". In: Phys. Rev. D 90.10, 103530 (Nov. 2014), p. 103530. DOI: `10.1103/PhysRevD.90.103530`. arXiv: `1408.1081` `[astro-ph.CO]`.

[253] Adrian E. Bayer et al. "Super-sample covariance of the power spectrum, bispectrum, halos, voids, and their cross-covariances". In: *arXiv e-prints*, arXiv:2210.15647 (Oct. 2022), arXiv:2210.15647. DOI: `10.48550/arXiv.2210.15647`. arXiv: `2210.15647` `[astro-ph.CO]`.

[254] Raul E. Angulo, Oliver Hahn, and Tom Abel. "How closely do baryons follow dark matter on large scales?" In: MNRAS 434.2 (Sept. 2013), pp. 1756–1764. DOI: 10.1093/mnras/stt1135. arXiv: 1301.7426 [astro-ph.CO].

[255] Simeon Bird et al. "More accurate simulations with separate initial conditions for baryons and dark matter". In: J. Cosmology Astropart. Phys. 2020.6, 002 (June 2020), p. 002. DOI: 10.1088/1475-7516/2020/06/002. arXiv: 2002.00015 [astro-ph.CO].

[256] Oliver Hahn, Cornelius Rampf, and Cora Uhlemann. "Higher order initial conditions for mixed baryon-CDM simulations". In: MNRAS 503.1 (May 2021), pp. 426–445. DOI: 10.1093/mnras/staa3773. arXiv: 2008.09124 [astro-ph.CO].

[257] Cora Dvorkin et al. "The Physics of Light Relics". In: *arXiv e-prints*, arXiv:2203.07943 (Mar. 2022), arXiv:2203.07943. arXiv: 2203.07943 [hep-ph].

[258] Weishuang Linda Xu, Julian B. Muñoz, and Cora Dvorkin. "Cosmological constraints on light but massive relics". In: Phys. Rev. D 105.9, 095029 (May 2022), p. 095029. DOI: 10.1103/PhysRevD.105.095029. arXiv: 2107.09664 [astro-ph.CO].

[259] Nicholas DePorzio et al. "Finding eV-scale light relics with cosmological observables". In: Phys. Rev. D 103.2, 023504 (Jan. 2021), p. 023504. DOI: 10.1103/PhysRevD.103.023504. arXiv: 2006.09380 [astro-ph.CO].

[260] Arka Banerjee et al. "Signatures of Light Massive Relics on non-linear structure formation". In: MNRAS 516.2 (Oct. 2022), pp. 2038–2049. DOI: 10.1093/mnras/stac2128. arXiv: 2202.09840 [astro-ph.CO].

[261] Pieter Daniel Meerburg et al. "Primordial Non-Gaussianity". In: BAAS 51.3, 107 (May 2019), p. 107. DOI: 10.48550/arXiv.1903.04409. arXiv: 1903.04409 [astro-ph.CO].

[262] James M. Bardeen. "Gauge-invariant cosmological perturbations". In: Phys. Rev. D 22.8 (Oct. 1980), pp. 1882–1905. DOI: 10.1103/PhysRevD.22.1882.

[263] Marcelo Alvarez et al. "Testing Inflation with Large Scale Structure: Connecting Hopes with Reality". In: *arXiv e-prints*, arXiv:1412.4671 (Dec. 2014), arXiv:1412.4671. DOI: 10.48550/arXiv.1412.4671. arXiv: 1412.4671 [astro-ph.CO].

[264] Ana Achúcarro et al. "Inflation: Theory and Observations". In: *arXiv e-prints*, arXiv:2203.08128 (Mar. 2022), arXiv:2203.08128. DOI: 10.48550/arXiv.2203.08128. arXiv: 2203.08128 [astro-ph.CO].

[265] Planck Collaboration et al. "Planck 2018 results. IX. Constraints on primordial non-Gaussianity". In: A&A 641, A9 (Sept. 2020), A9. DOI: 10.1051/0004-6361/201935891. arXiv: 1905.05697 [astro-ph.CO].

[266] 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].

[267] Peter Ade et al. "The Simons Observatory: science goals and forecasts". In: J. Cosmology Astropart. Phys. 2019.2, 056 (Feb. 2019), p. 056. DOI: 10.1088/1475-7516/2019/02/056. arXiv: 1808.07445 [astro-ph.CO].

[268] Roland de Putter and Olivier Doré. "Designing an inflation galaxy survey: How to measure σ (f_{NL})$\sim$1 using scale-dependent galaxy bias". In: Phys. Rev. D 95.12, 123513 (June 2017), p. 123513. DOI: 10.1103/PhysRevD.95.123513. arXiv: 1412.3854 [astro-ph.CO].

[269] Noah Sailer et al. "Cosmology at high redshift - a probe of fundamental physics". In: J. Cosmology Astropart. Phys. 2021.12, 049 (Dec. 2021), p. 049. DOI: 10.1088/1475-7516/2021/12/049. arXiv: 2106.09713 [astro-ph.CO].

[270] Marcel Schmittfull and Uroš Seljak. "Parameter constraints from cross-correlation of CMB lensing with galaxy clustering". In: Phys. Rev. D 97.12, 123540 (June 2018), p. 123540. DOI: 10.1103/PhysRevD.97.123540. arXiv: 1710.09465 [astro-ph.CO].

[271] Utkarsh Giri, Moritz Münchmeyer, and Kendrick M. Smith. "Robust Neural Network-Enhanced Estimation of Local Primordial Non-Gaussianity". In: *arXiv e-prints*, arXiv:2205.12964 (May 2022), arXiv:2205.12964. arXiv: 2205.12964 [astro-ph.CO].

[272] Azadeh Moradinezhad Dizgah et al. "Primordial non-Gaussianity from biased tracers: likelihood analysis of real-space power spectrum and bispectrum". In: J. Cosmology Astropart. Phys. 2021.5, 015 (May 2021), p. 015. DOI: 10.1088/1475-7516/2021/05/015. arXiv: 2010.14523 [astro-ph.CO].

[273] L. Raul Abramo, João Vitor Dinarte Ferri, and Ian Lucas Tashiro. "Fisher matrix for multiple tracers: the information in the cross-spectra". In: J. Cosmology Astropart. Phys. 2022.4, 013 (Apr. 2022), p. 013. DOI: 10.1088/1475-7516/2022/04/013. arXiv: 2112.01812 [astro-ph.CO].

[274] Sheean Jolicoeur, Roy Maartens, and Simthembile Dlamini. "Constraining primordial non-Gaussianity by combining next-generation galaxy and 21 cm intensity mapping surveys". In: *arXiv e-prints*, arXiv:2301.02406 (Jan. 2023), arXiv:2301.02406. DOI: 10.48550/arXiv.2301.02406. arXiv: 2301.02406 [astro-ph.CO].

[275] Uroš Seljak. "Extracting Primordial Non-Gaussianity without Cosmic Variance". In: Phys. Rev. Lett. 102.2, 021302 (Jan. 2009), p. 021302. DOI: 10.1103/PhysRevLett.102.021302. arXiv: 0807.1770 [astro-ph].

[276] Patrick McDonald and Uroš Seljak. "How to evade the sample variance limit on measurements of redshift-space distortions". In: J. Cosmology Astropart. Phys. 2009.10, 007 (Oct. 2009), p. 007. DOI: 10.1088/1475-7516/2009/10/007. arXiv: 0810.0323 [astro-ph].

[277] Simone Ferraro and Kendrick M. Smith. "Using large scale structure to measure f_{NL} , g_{NL} and τ_{NL}". In: Phys. Rev. D 91.4, 043506 (Feb. 2015), p. 043506. DOI: 10.1103/PhysRevD.91.043506. arXiv: 1408.3126 [astro-ph.CO].

[278] Gary M. Bernstein and Yan-Chuan Cai. "Cosmology without cosmic variance". In: MNRAS 416.4 (Oct. 2011), pp. 3009–3016. DOI: `10.1111/j.1365-2966.2011.19249.x`. arXiv: `1104.3862 [astro-ph.CO]`.

[279] Omar Darwish et al. "Density reconstruction from biased tracers and its application to primordial non-Gaussianity". In: Phys. Rev. D 104.12, 123520 (Dec. 2021), p. 123520. DOI: `10.1103/PhysRevD.104.123520`. arXiv: `2007.08472 [astro-ph.CO]`.

[280] Thiago Mergulhão et al. "The effective field theory of large-scale structure and multi-tracer". In: J. Cosmology Astropart. Phys. 2022.4, 021 (Apr. 2022), p. 021. DOI: `10.1088/1475-7516/2022/04/021`. arXiv: `2108.11363 [astro-ph.CO]`.

[281] Yuting Wang et al. "The clustering of the SDSS-IV extended baryon oscillation spectroscopic survey DR16 luminous red galaxy and emission-line galaxy samples: cosmic distance and structure growth measurements using multiple tracers in configuration space". In: MNRAS 498.3 (Nov. 2020), pp. 3470–3483. DOI: `10.1093/mnras/staa2593`. arXiv: `2007.09010 [astro-ph.CO]`.

[282] Yuting Wang and Gong-Bo Zhao. "A brief review on cosmological analysis of galaxy surveys with multiple tracers". In: *Research in Astronomy and Astrophysics* 20.10, 158 (Oct. 2020), p. 158. DOI: `10.1088/1674-4527/20/10/158`. arXiv: `2009.03862 [astro-ph.CO]`.

[283] Neal Dalal et al. "Imprints of primordial non-Gaussianities on large-scale structure: Scale-dependent bias and abundance of virialized objects". In: Phys. Rev. D 77.12, 123514 (June 2008), p. 123514. DOI: `10.1103/PhysRevD.77.123514`. arXiv: `0710.4560 [astro-ph]`.

[284] Anže Slosar et al. "Constraints on local primordial non-Gaussianity from large scale structure". In: J. Cosmology Astropart. Phys. 2008.8, 031 (Aug. 2008), p. 031. DOI: `10.1088/1475-7516/2008/08/031`. arXiv: `0805.3580 [astro-ph]`.

[285] Boris Leistedt, Hiranya V. Peiris, and Nina Roth. "Constraints on Primordial Non-Gaussianity from 800 000 Photometric Quasars". In: Phys. Rev. Lett. 113.22, 221301 (Nov. 2014), p. 221301. DOI: `10.1103/PhysRevLett.113.221301`. arXiv: `1405.4315 [astro-ph.CO]`.

[286] Eva-Maria Mueller et al. "The clustering of galaxies in the completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Primordial non-Gaussianity in Fourier Space". In: *arXiv e-prints*, arXiv:2106.13725 (June 2021), arXiv:2106.13725. arXiv: `2106.13725 [astro-ph.CO]`.

[287] Emanuele Castorina et al. "Redshift-weighted constraints on primordial non-Gaussianity from the clustering of the eBOSS DR14 quasars in Fourier space". In: J. Cosmology Astropart. Phys. 2019.9, 010 (Sept. 2019), p. 010. DOI: `10.1088/1475-7516/2019/09/010`. arXiv: `1904.08859 [astro-ph.CO]`.

[288] Giovanni Cabass et al. "Constraints on multifield inflation from the BOSS galaxy survey". In: Phys. Rev. D 106.4, 043506 (Aug. 2022), p. 043506. DOI: `10.1103/PhysRevD.106.043506`. arXiv: `2204.01781 [astro-ph.CO]`.

[289] Guido D'Amico et al. "Limits on primordial non-Gaussianities from BOSS galaxy-clustering data". In: *arXiv e-prints*, arXiv:2201.11518 (Jan. 2022), arXiv:2201.11518. arXiv: `2201.11518 [astro-ph.CO]`.

[290] Alexandre Barreira. "Can we actually constrain f_{NL} using the scale-dependent bias effect? An illustration of the impact of galaxy bias uncertainties using the BOSS DR12 galaxy power spectrum". In: J. Cosmology Astropart. Phys. 2022.11, 013 (Nov. 2022), p. 013. DOI: `10.1088/1475-7516/2022/11/013`. arXiv: `2205.05673 [astro-ph.CO]`.

[291] Alexandre Barreira et al. "Galaxy bias and primordial non-Gaussianity: insights from galaxy formation simulations with IllustrisTNG". In: J. Cosmology Astropart. Phys. 2020.12, 013 (Dec. 2020), p. 013. DOI: `10.1088/1475-7516/2020/12/013`. arXiv: `2006.09368 [astro-ph.CO]`.

[292] Alexandre Barreira. "On the impact of galaxy bias uncertainties on primordial non-Gaussianity constraints". In: J. Cosmology Astropart. Phys. 2020.12, 031 (Dec. 2020), p. 031. DOI: `10.1088/1475-7516/2020/12/031`. arXiv: `2009.06622 [astro-ph.CO]`.

[293] Alexandre Barreira. "Predictions for local PNG bias in the galaxy power spectrum and bispectrum and the consequences for f_{NL} constraints". In: J. Cosmology Astropart. Phys. 2022.1, 033 (Jan. 2022), p. 033. DOI: `10.1088/1475-7516/2022/01/033`. arXiv: `2107.06887 [astro-ph.CO]`.

[294] M. Marinucci, V. Desjacques, and A. Benson. "Non-Gaussian assembly bias from a semi-analytic galaxy formation model". In: MNRAS 524.1 (Sept. 2023), pp. 325–337. DOI: `10.1093/mnras/stad1884`. arXiv: `2303.10337 [astro-ph.CO]`.

[295] Beth A. Reid et al. "Non-Gaussian halo assembly bias". In: J. Cosmology Astropart. Phys. 2010.7, 013 (July 2010), p. 013. DOI: `10.1088/1475-7516/2010/07/013`. arXiv: `1004.1637 [astro-ph.CO]`.

[296] Titouan Lazeyras et al. "Assembly bias in the local PNG halo bias and its implication for $f_{\rm NL}$ constraints". In: *arXiv e-prints*, arXiv:2209.07251 (Sept. 2022), arXiv:2209.07251. arXiv: `2209.07251 [astro-ph.CO]`.

[297] Matteo Biagetti et al. "Verifying the consistency relation for the scale-dependent bias from local primordial non-Gaussianity". In: MNRAS 468.3 (July 2017), pp. 3277–3288. DOI: `10.1093/mnras/stx714`. arXiv: `1611.04901 [astro-ph.CO]`.

[298] Tobias Baldauf et al. "Linear response to long wavelength fluctuations using curvature simulations". In: J. Cosmology Astropart. Phys. 2016.9, 007 (Sept. 2016), p. 007. DOI: `10.1088/1475-7516/2016/09/007`. arXiv: `1511.01465 [astro-ph.CO]`.

[299] Edwin Sirko. "Initial Conditions to Cosmological N-Body Simulations, or, How to Run an Ensemble of Simulations". In: ApJ 634.2 (Nov. 2005), pp. 728–743. DOI: `10.1086/497090`. arXiv: `astro-ph/0503106 [astro-ph]`.

[300] Nickolay Y. Gnedin, Andrey V. Kravtsov, and Douglas H. Rudd. "Implementing the DC Mode in Cosmological Simulations with Supercomoving Variables". In: ApJS 194.2, 46 (June 2011), p. 46. DOI: `10.1088/0067-0049/194/2/46`. arXiv: `1104.1428` `[astro-ph.CO]`.

[301] C. Wagner et al. "Separate universe simulations." In: MNRAS 448 (Mar. 2015), pp. L11–L15. DOI: `10.1093/mnrasl/slu187`. arXiv: `1409.6294 [astro-ph.CO]`.

[302] Tobias Baldauf et al. "Galaxy bias and non-linear structure formation in general relativity". In: J. Cosmology Astropart. Phys. 2011.10, 031 (Oct. 2011), p. 031. DOI: `10.1088/1475-7516/2011/10/031`. arXiv: `1106.5507 [astro-ph.CO]`.

[303] Yin Li, Wayne Hu, and Masahiro Takada. "Super-sample covariance in simulations". In: Phys. Rev. D 89.8, 083519 (Apr. 2014), p. 083519. DOI: `10.1103/PhysRevD.89.083519`. arXiv: `1401.0385 [astro-ph.CO]`.

[304] Patrick McDonald. "Toward a Measurement of the Cosmological Geometry at z ~2: Predicting Lyα Forest Correlation in Three Dimensions and the Potential of Future Data Sets". In: ApJ 585.1 (Mar. 2003), pp. 34–51. DOI: `10.1086/345945`. arXiv: `astro-ph/0108064 [astro-ph]`.

[305] Liang Dai, Enrico Pajer, and Fabian Schmidt. "On separate universes". In: J. Cosmology Astropart. Phys. 2015.10 (Oct. 2015), pp. 059–059. DOI: `10.1088/1475-7516/2015/10/059`. arXiv: `1504.00351 [astro-ph.CO]`.

[306] Dylan Nelson et al. "The IllustrisTNG simulations: public data release". In: *Computational Astrophysics and Cosmology* 6.1, 2 (May 2019), p. 2. DOI: `10.1186/s40668-019-0028-x`. arXiv: `1812.05609 [astro-ph.GA]`.

[307] Annalisa Pillepich et al. "First results from the IllustrisTNG simulations: the stellar mass content of groups and clusters of galaxies". In: MNRAS 475.1 (Mar. 2018), pp. 648–675. DOI: `10.1093/mnras/stx3112`. arXiv: `1707.03406 [astro-ph.GA]`.

[308] Volker Springel et al. "First results from the IllustrisTNG simulations: matter and galaxy clustering". In: MNRAS 475.1 (Mar. 2018), pp. 676–698. DOI: `10.1093/mnras/stx3304`. arXiv: `1707.03397 [astro-ph.GA]`.

[309] Dylan Nelson et al. "First results from the IllustrisTNG simulations: the galaxy colour bimodality". In: MNRAS 475.1 (Mar. 2018), pp. 624–647. DOI: `10.1093/mnras/stx3040`. arXiv: `1707.03395 [astro-ph.GA]`.

[310] Jill P. Naiman et al. "First results from the IllustrisTNG simulations: a tale of two elements - chemical evolution of magnesium and europium". In: MNRAS 477.1 (June 2018), pp. 1206–1224. DOI: `10.1093/mnras/sty618`. arXiv: `1707.03401 [astro-ph.GA]`.

[311] Federico Marinacci et al. "First results from the IllustrisTNG simulations: radio haloes and magnetic fields". In: MNRAS 480.4 (Nov. 2018), pp. 5113–5139. DOI: `10.1093/mnras/sty2206`. arXiv: `1707.03396 [astro-ph.CO]`.

[312] 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].

[313] M. Davis et al. "The evolution of large-scale structure in a universe dominated by cold dark matter". In: ApJ 292 (May 1985), pp. 371–394. DOI: 10.1086/163168.

[314] Dhayaa Anbajagane, August E. Evrard, and Arya Farahi. "Baryonic imprints on DM haloes: population statistics from dwarf galaxies to galaxy clusters". In: MNRAS 509.3 (Jan. 2022), pp. 3441–3461. DOI: 10.1093/mnras/stab3177. arXiv: 2109.02713 [astro-ph.CO].

[315] Julio F. Navarro, Carlos S. Frenk, and Simon D. M. White. "A Universal Density Profile from Hierarchical Clustering". In: ApJ 490.2 (Dec. 1997), pp. 493–508. DOI: 10.1086/304888. arXiv: astro-ph/9611107 [astro-ph].

[316] Boryana Hadzhiyska et al. "The galaxy-halo connection of emission-line galaxies in IllustrisTNG". In: MNRAS 502.3 (Apr. 2021), pp. 3599–3617. DOI: 10.1093/mnras/stab243. arXiv: 2011.05331 [astro-ph.GA].

[317] Sihan Yuan et al. "Illustrating galaxy-halo connection in the DESI era with ILLUS-TRISTNG". In: MNRAS 512.4 (June 2022), pp. 5793–5811. DOI: 10.1093/mnras/stac830. arXiv: 2202.12911 [astro-ph.CO].

[318] Boryana Hadzhiyska et al. "The MillenniumTNG Project: Refining the one-halo model of red and blue galaxies at different redshifts". In: *arXiv e-prints*, arXiv:2210.10068 (Oct. 2022), arXiv:2210.10068. DOI: 10.48550/arXiv.2210.10068. arXiv: 2210.10068 [astro-ph.CO].

[319] Rongpu Zhou et al. "Target Selection and Validation of DESI Luminous Red Galaxies". In: AJ 165.2, 58 (Feb. 2023), p. 58. DOI: 10.3847/1538-3881/aca5fb. arXiv: 2208.08515 [astro-ph.CO].

[320] A. Raichoor et al. "Target Selection and Validation of DESI Emission Line Galaxies". In: *arXiv e-prints*, arXiv:2208.08513 (Aug. 2022), arXiv:2208.08513. arXiv: 2208.08513 [astro-ph.CO].

[321] Rongpu Zhou et al. "The clustering of DESI-like luminous red galaxies using photometric redshifts". In: MNRAS 501.3 (Mar. 2021), pp. 3309–3331. DOI: 10.1093/mnras/staa3764. arXiv: 2001.06018 [astro-ph.CO].

[322] Pablo Villanueva-Domingo et al. "Inferring Halo Masses with Graph Neural Networks". In: ApJ 935.1, 30 (Aug. 2022), p. 30. DOI: 10.3847/1538-4357/ac7aa3. arXiv: 2111.08683 [astro-ph.CO].

[323] Shankar Agarwal, Romeel Davé, and Bruce A. Bassett. "Painting galaxies into dark matter haloes using machine learning". In: MNRAS 478.3 (Aug. 2018), pp. 3410–3422. DOI: 10.1093/mnras/sty1169. arXiv: 1712.03255 [astro-ph.GA].

[324] Benjamin P. Moster et al. "GalaxyNet: connecting galaxies and dark matter haloes with deep neural networks and reinforcement learning in large volumes". In: MNRAS 507.2 (Oct. 2021), pp. 2115–2136. DOI: 10.1093/mnras/stab1449. arXiv: 2005.12276 [astro-ph.GA].

[325] Xiaoju Xu et al. "Predicting halo occupation and galaxy assembly bias with machine learning". In: MNRAS 507.4 (Nov. 2021), pp. 4879–4899. DOI: 10.1093/mnras/stab2464. arXiv: 2107.01223 [astro-ph.CO].

[326] Ana Maria Delgado et al. "Modelling the galaxy-halo connection with machine learning". In: MNRAS 515.2 (Sept. 2022), pp. 2733–2746. DOI: 10.1093/mnras/stac1951. arXiv: 2111.02422 [astro-ph.CO].

[327] Natalí S. M. de Santi et al. "Mimicking the halo-galaxy connection using machine learning". In: MNRAS 514.2 (Aug. 2022), pp. 2463–2478. DOI: 10.1093/mnras/stac1469. arXiv: 2201.06054 [astro-ph.GA].

[328] Richard Stiskalek et al. "The scatter in the galaxy-halo connection: a machine learning analysis". In: MNRAS 514.3 (Aug. 2022), pp. 4026–4045. DOI: 10.1093/mnras/stac1609. arXiv: 2202.14006 [astro-ph.GA].

[329] Tianqi Chen and Carlos Guestrin. "XGBoost: A Scalable Tree Boosting System". In: *arXiv e-prints*, arXiv:1603.02754 (Mar. 2016), arXiv:1603.02754. DOI: 10.48550/arXiv.1603.02754. arXiv: 1603.02754 [cs.LG].

[330] Scott Lundberg and Su-In Lee. "A Unified Approach to Interpreting Model Predictions". In: *arXiv e-prints*, arXiv:1705.07874 (May 2017), arXiv:1705.07874. DOI: 10.48550/arXiv.1705.07874. arXiv: 1705.07874 [cs.AI].

[331] Simone Ferraro and Michael J. Wilson. "Inflation and Dark Energy from spectroscopy at z ¿ 2". In: BAAS 51.3, 72 (May 2019), p. 72. DOI: 10.48550/arXiv.1903.09208. arXiv: 1903.09208 [astro-ph.CO].

[332] 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].

[333] Planck Collaboration et al. "Planck 2015 results. XIII. Cosmological parameters". In: A&A 594, A13 (Sept. 2016), A13. DOI: 10.1051/0004-6361/201525830. arXiv: 1502.01589 [astro-ph.CO].

[334] Volker Springel et al. "First results from the IllustrisTNG simulations: matter and galaxy clustering". In: MNRAS 475.1 (Mar. 2018), pp. 676–698. DOI: 10.1093/mnras/stx3304. arXiv: 1707.03397 [astro-ph.GA].

[335] Nico Hamaus, Uroš Seljak, and Vincent Desjacques. "Optimal constraints on local primordial non-Gaussianity from the two-point statistics of large-scale structure". In: Phys. Rev. D 84.8, 083509 (Oct. 2011), p. 083509. DOI: 10.1103/PhysRevD.84.083509. arXiv: 1104.2321 [astro-ph.CO].

[336] C. Alcock and B. Paczynski. "An evolution free test for non-zero cosmological constant". In: Nature 281 (Oct. 1979), p. 358. DOI: 10.1038/281358a0.

[337] Antony Lewis, Anthony Challinor, and Anthony Lasenby. "Efficient Computation of Cosmic Microwave Background Anisotropies in Closed Friedmann-Robertson-Walker Models". In: ApJ 538.2 (Aug. 2000), pp. 473–476. DOI: 10.1086/309179. arXiv: astro-ph/9911177 [astro-ph].

[338] Rongpu Zhou et al. "Target Selection and Validation of DESI Luminous Red Galaxies". In: AJ 165.2, 58 (Feb. 2023), p. 58. DOI: 10.3847/1538-3881/aca5fb. arXiv: 2208.08515 [astro-ph.CO].

[339] Nick Mostek et al. "The DEEP2 Galaxy Redshift Survey: Clustering Dependence on Galaxy Stellar Mass and Star Formation Rate at z ~1". In: ApJ 767.1, 89 (Apr. 2013), p. 89. DOI: 10.1088/0004-637X/767/1/89. arXiv: 1210.6694 [astro-ph.CO].

[340] Giovanni Cabass et al. "Constraining Single-Field Inflation with MegaMapper". In: *arXiv e-prints*, arXiv:2211.14899 (Nov. 2022), arXiv:2211.14899. arXiv: 2211.14899 [astro-ph.CO].

[341] M. J. Wilson and Martin White. "Cosmology with dropout selection: straw-man surveys & CMB lensing". In: J. Cosmology Astropart. Phys. 2019.10, 015 (Oct. 2019), p. 015. DOI: 10.1088/1475-7516/2019/10/015. arXiv: 1904.13378 [astro-ph.CO].

[342] Suman Bhattacharya et al. "Mass Function Predictions Beyond ΛCDM". In: ApJ 732.2, 122 (May 2011), p. 122. DOI: 10.1088/0004-637X/732/2/122. arXiv: 1005.2239 [astro-ph.CO].

[343] Benedikt Diemer. "COLOSSUS: A Python Toolkit for Cosmology, Large-scale Structure, and Dark Matter Halos". In: ApJS 239.2, 35 (Dec. 2018), p. 35. DOI: 10.3847/1538-4365/aaee8c. arXiv: 1712.04512 [astro-ph.CO].

[344] Jeremy Tinker et al. "Toward a Halo Mass Function for Precision Cosmology: The Limits of Universality". In: ApJ 688.2 (Dec. 2008), pp. 709–728. DOI: 10.1086/591439. arXiv: 0803.2706 [astro-ph].

[345] Jeremy L. Tinker et al. "The Large-scale Bias of Dark Matter Halos: Numerical Calibration and Model Tests". In: ApJ 724.2 (Dec. 2010), pp. 878–886. DOI: 10.1088/0004-637X/724/2/878. arXiv: 1001.3162 [astro-ph.CO].

[346] Max Tegmark, Andy N. Taylor, and Alan F. Heavens. "Karhunen-Loève Eigenvalue Problems in Cosmology: How Should We Tackle Large Data Sets?" In: ApJ 480.1 (May 1997), pp. 22–35. DOI: 10.1086/303939. arXiv: astro-ph/9603021 [astro-ph].

[347] Tobias Baldauf et al. "Halo stochasticity from exclusion and nonlinear clustering". In: Phys. Rev. D 88.8, 083507 (Oct. 2013), p. 083507. DOI: 10.1103/PhysRevD.88.083507. arXiv: 1305.2917 [astro-ph.CO].

[348] Dimitry Ginzburg and Vincent Desjacques. "Shot noise in multitracer constraints on f_{NL} and relativistic projections: Power spectrum". In: MNRAS 495.1 (June 2020), pp. 932–942. DOI: 10.1093/mnras/staa1154. arXiv: 1911.11701 [astro-ph.CO].

[349] Anže Slosar. "Optimal weighting in f_{NL} constraints from large scale structure in an idealised case". In: J. Cosmology Astropart. Phys. 2009.3, 004 (Mar. 2009), p. 004. DOI: 10.1088/1475-7516/2009/03/004. arXiv: 0808.0044 [astro-ph].

[350] Emanuele Castorina et al. "Primordial Non-Gaussianities and Zero-Bias Tracers of the Large-Scale Structure". In: Phys. Rev. Lett. 121.10, 101301 (Sept. 2018), p. 101301. DOI: 10.1103/PhysRevLett.121.101301. arXiv: 1803.11539 [astro-ph.CO].

[351] Ruediger Pakmor et al. "The MillenniumTNG Project: The hydrodynamical full physics simulation and a first look at its galaxy clusters". In: *arXiv e-prints*, arXiv:2210.10060 (Oct. 2022), arXiv:2210.10060. DOI: 10.48550/arXiv.2210.10060. arXiv: 2210.10060 [astro-ph.CO].

[352] Daniel J. Eisenstein et al. "Improving Cosmological Distance Measurements by Reconstruction of the Baryon Acoustic Peak". In: ApJ 664.2 (Aug. 2007), pp. 675–679. DOI: 10.1086/518712. arXiv: astro-ph/0604362 [astro-ph].

[353] Rodrigo Voivodic and Alexandre Barreira. "Responses of Halo Occupation Distributions: a new ingredient in the halo model & the impact on galaxy bias". In: J. Cosmology Astropart. Phys. 2021.5, 069 (May 2021), p. 069. DOI: 10.1088/1475-7516/2021/05/069. arXiv: 2012.04637 [astro-ph.CO].

[354] Alexandre Barreira and Elisabeth Krause. "Towards optimal and robust f_{NL} constraints with multi-tracer analyses". In: *arXiv e-prints*, arXiv:2302.09066 (Feb. 2023), arXiv:2302.09066. arXiv: 2302.09066 [astro-ph.CO].

[355] Daisuke Yamauchi, Shuichiro Yokoyama, and Keitaro Takahashi. "Multitracer technique for galaxy bispectrum: An application to constraints on nonlocal primordial non-Gaussianities". In: Phys. Rev. D 95.6, 063530 (Mar. 2017), p. 063530. DOI: 10.1103/PhysRevD.95.063530. arXiv: 1611.03590 [astro-ph.CO].

[356] Jeffrey Larson, Matt Menickelly, and Stefan M. Wild. "Derivative-free optimization methods". In: *Acta Numerica* 28 (2019), pp. 287–404. DOI: 10.1017/S0962492919000060.

[357] K.V. Price, R.N. Storn, and J.A. Lampinen. *Differential Evolution: A Practical Approach to Global Optimization*. Natural Computing Series. 2005.

[358] Jorge Nocedal and Stephen J. Wright. *Numerical Optimization*. 2e. New York, NY, USA: Springer, 2006.

[359] Kenneth A. De Jong. "Evolutionary computation: a unified approach". In: (2019). Ed. by Manuel López-Ibáñez, Anne Auger, and Thomas Stützle, pp. 507–522. DOI: 10.1145/3319619.3323379. URL: https://doi.org/10.1145/3319619.3323379.

[360] J. A. Nelder and R. Mead. "A Simplex Method for Function Minimization". In: *The Computer Journal* 7.4 (Jan. 1965), pp. 308–313. ISSN: 0010-4620. DOI: 10.1093/comjnl/7.4.308. eprint: https://academic.oup.com/comjnl/article-pdf/7/4/308/1013182/7-4-308.pdf. URL: https://doi.org/10.1093/comjnl/7.4.308.

[361] Burr Settles. "From Theories to Queries: Active Learning in Practice". In: *Active Learning and Experimental Design workshop In conjunction with AISTATS 2010*. Ed. by Isabelle Guyon et al. Vol. 16. Proceedings of Machine Learning Research. Sardinia, Italy: PMLR, 16 May 2011, pp. 1–18. URL: https://proceedings.mlr.press/v16/settles11a.html.

[362] Bobak Shahriari et al. "Taking the Human Out of the Loop: A Review of Bayesian Optimization". eng. In: *Proceedings of the IEEE* 104.1 (2016), pp. 148–175. ISSN: 0018-9219.

[363] Eric Brochu, Vlad M. Cora, and Nando de Freitas. "A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning". In: *arXiv e-prints*, arXiv:1012.2599 (Dec. 2010), arXiv:1012.2599. arXiv: 1012.2599 [cs.LG].

[364] Jasper Snoek, Hugo Larochelle, and Ryan P. Adams. "Practical Bayesian Optimization of Machine Learning Algorithms". In: *arXiv e-prints*, arXiv:1206.2944 (June 2012), arXiv:1206.2944. arXiv: 1206.2944 [stat.ML].

[365] Roman Garnett. *Bayesian Optimization*. Cambridge University Press, 2023.

[366] Samuel Müller et al. "PFNs4BO: In-Context Learning for Bayesian Optimization". In: *arXiv e-prints*, arXiv:2305.17535 (May 2023), arXiv:2305.17535. DOI: 10.48550/arXiv.2305.17535. arXiv: 2305.17535 [cs.LG].

[367] Hugh Chipman, Pritam Ranjan, and Weiwei Wang. "Sequential design for computer experiments with a flexible Bayesian additive model". In: *Canadian Journal of Statistics* 40.4 (2012), pp. 663–678. DOI: https://doi.org/10.1002/cjs.11156. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cjs.11156. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/cjs.11156.

[368] Malek Ben Salem et al. "Universal Prediction Distribution for Surrogate Models". In: *SIAM/ASA Journal on Uncertainty Quantification* 5.1 (2017), pp. 1086–1109. DOI: 10.1137/15M1053529. eprint: https://doi.org/10.1137/15M1053529. URL: https://doi.org/10.1137/15M1053529.

[369] Peter I. Frazier. "A Tutorial on Bayesian Optimization". In: *arXiv e-prints*, arXiv:1807.02811 (July 2018), arXiv:1807.02811. arXiv: 1807.02811 [stat.ML].

[370] Xiaoyu Lu, Alexis Boukouvalas, and James Hensman. "Additive Gaussian Processes Revisited". In: *Proceedings of the 39th International Conference on Machine Learning*. Ed. by Kamalika Chaudhuri et al. Vol. 162. Proceedings of Machine Learning Research. PMLR, 17–23 Jul 2022, pp. 14358–14383. URL: https://proceedings.mlr.press/v162/lu22b.html.

[371] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. 2006.

[372] Mickael Binois and Nathan Wycoff. "A survey on high-dimensional Gaussian process modeling with application to Bayesian optimization". In: *arXiv e-prints*, arXiv:2111.05040 (Nov. 2021), arXiv:2111.05040. arXiv: `2111.05040 [math.OC]`.

[373] Jasper Snoek et al. "Scalable Bayesian Optimization Using Deep Neural Networks". In: *arXiv e-prints*, arXiv:1502.05700 (Feb. 2015), arXiv:1502.05700. arXiv: `1502.05700 [stat.ML]`.

[374] Jost Tobias Springenberg et al. "Bayesian Optimization with Robust Bayesian Neural Networks". In: *Advances in Neural Information Processing Systems*. Ed. by D. Lee et al. Vol. 29. Curran Associates, Inc., 2016. URL: `https://proceedings.neurips.cc/paper/2016/file/a96d3afec184766bfeca7a9f989fc7e7-Paper.pdf`.

[375] Frank Hutter, Holger H. Hoos, and Kevin Leyton-Brown. "Sequential Model-Based Optimization for General Algorithm Configuration". In: *Learning and Intelligent Optimization*. Ed. by Carlos A. Coello Coello. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 507–523. ISBN: 978-3-642-25566-3.

[376] Rommel G. Regis and Christine A. Shoemaker. "Combining radial basis function surrogates and dynamic coordinate search in high-dimensional expensive black-box optimization". In: *Engineering Optimization* 45.5 (2013), pp. 529–555. DOI: `10.1080/0305215X.2012.687731`. eprint: `https://doi.org/10.1080/0305215X.2012.687731`. URL: `https://doi.org/10.1080/0305215X.2012.687731`.

[377] Florian Wenzel et al. "Hyperparameter Ensembles for Robustness and Uncertainty Quantification". In: *arXiv e-prints*, arXiv:2006.13570 (June 2020), arXiv:2006.13570. arXiv: `2006.13570 [cs.LG]`.

[378] Jakob Heiss et al. "NOMU: Neural Optimization-based Model Uncertainty". In: *arXiv e-prints*, arXiv:2102.13640 (Feb. 2021), arXiv:2102.13640. arXiv: `2102.13640 [cs.LG]`.

[379] Arsenii Ashukha et al. "Pitfalls of In-Domain Uncertainty Estimation and Ensembling in Deep Learning". In: *arXiv e-prints*, arXiv:2002.06470 (Feb. 2020), arXiv:2002.06470. arXiv: `2002.06470 [stat.ML]`.

[380] Matthias Minderer et al. "Revisiting the Calibration of Modern Neural Networks". In: *arXiv e-prints*, arXiv:2106.07998 (June 2021), arXiv:2106.07998. arXiv: `2106.07998 [cs.LG]`.

[381] David Eriksson et al. "Scalable Global Optimization via Local Bayesian Optimization". In: *arXiv e-prints*, arXiv:1910.01739 (Oct. 2019), arXiv:1910.01739. arXiv: `1910.01739 [cs.LG]`.

[382] Richard D. P. Grumitt, Biwei Dai, and Uros Seljak. "Deterministic Langevin Monte Carlo with Normalizing Flows for Bayesian Inference". In: *arXiv e-prints*, arXiv:2205.14240 (May 2022), arXiv:2205.14240. arXiv: `2205.14240 [stat.ML]`.

[383] Dimitra Maoutsa, Sebastian Reich, and Manfred Opper. "Interacting Particle Solutions of Fokker-Planck Equations Through Gradient-Log-Density Estimation". In: *Entropy* 22.8 (July 2020), p. 802. DOI: 10.3390/e22080802. arXiv: 2006.00702 [cond-mat.stat-mech].

[384] Yang Song et al. "Score-based generative modeling through stochastic differential equations". In: *arXiv preprint arXiv:2011.13456* (2020).

[385] Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. "Density estimation using Real NVP". In: *arXiv e-prints*, arXiv:1605.08803 (May 2016), arXiv:1605.08803. arXiv: 1605.08803 [cs.LG].

[386] George Papamakarios, Theo Pavlakou, and Iain Murray. "Masked Autoregressive Flow for Density Estimation". In: *arXiv e-prints*, arXiv:1705.07057 (May 2017), arXiv:1705.07057. DOI: 10.48550/arXiv.1705.07057. arXiv: 1705.07057 [stat.ML].

[387] Diederik P. Kingma and Prafulla Dhariwal. "Glow: Generative Flow with Invertible 1x1 Convolutions". In: *arXiv e-prints*, arXiv:1807.03039 (July 2018), arXiv:1807.03039. DOI: 10.48550/arXiv.1807.03039. arXiv: 1807.03039 [stat.ML].

[388] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. "Optimization by Simulated Annealing". In: *Science* 220.4598 (May 1983), pp. 671–680. DOI: 10.1126/science.220.4598.671.

[389] Arnaud Doucet, Nando de Freitas, and Neil Gordon. "An Introduction to Sequential Monte Carlo Methods". In: *Sequential Monte Carlo Methods in Practice*. Ed. by Arnaud Doucet, Nando de Freitas, and Neil Gordon. New York, NY: Springer New York, 2001, pp. 3–14.

[390] Nikolaus Hansen. "The CMA Evolution Strategy: A Tutorial". In: *arXiv e-prints*, arXiv:1604.00772 (Apr. 2016), arXiv:1604.00772. arXiv: 1604.00772 [cs.LG].

[391] B. P. Abbott et al. "Properties of the Binary Black Hole Merger GW150914". In: Phys. Rev. Lett. 116.24, 241102 (June 2016), p. 241102. DOI: 10.1103/PhysRevLett.116.241102. arXiv: 1602.03840 [gr-qc].

[392] W. J. Handley, M. P. Hobson, and A. N. Lasenby. "polychord: nested sampling for cosmology." In: MNRAS 450 (June 2015), pp. L61–L65. DOI: 10.1093/mnrasl/slv047. arXiv: 1502.01856 [astro-ph.CO].

[393] Beth A. Reid et al. "Cosmological constraints from the clustering of the Sloan Digital Sky Survey DR7 luminous red galaxies". In: MNRAS 404.1 (May 2010), pp. 60–85. DOI: 10.1111/j.1365-2966.2010.16276.x. arXiv: 0907.1659 [astro-ph.CO].

[394] Lukas Biewald. *Experiment Tracking with Weights and Biases*. Software available from wandb.com. 2020. URL: https://www.wandb.com/.

[395] Takuya Akiba et al. "Optuna: A Next-generation Hyperparameter Optimization Framework". In: *arXiv e-prints*, arXiv:1907.10902 (July 2019), arXiv:1907.10902. arXiv: `1907.10902 [cs.LG]`.

[396] Katharina Eggensperger et al. "HPOBench: A Collection of Reproducible Multi-Fidelity Benchmark Problems for HPO". In: *arXiv e-prints*, arXiv:2109.06716 (Sept. 2021), arXiv:2109.06716. arXiv: `2109.06716 [cs.LG]`.

[397] Marius Lindauer et al. "SMAC3: A Versatile Bayesian Optimization Package for Hyperparameter Optimization". In: *arXiv e-prints*, arXiv:2109.09831 (Sept. 2021), arXiv:2109.09831. arXiv: `2109.09831 [cs.LG]`.

[398] Ryan Turner et al. "Bayesian Optimization is Superior to Random Search for Machine Learning Hyperparameter Tuning: Analysis of the Black-Box Optimization Challenge 2020". In: *arXiv e-prints*, arXiv:2104.10201 (Apr. 2021), arXiv:2104.10201. arXiv: `2104.10201 [cs.LG]`.

[399] Colin White, Willie Neiswanger, and Yash Savani. "BANANAS: Bayesian Optimization with Neural Architectures for Neural Architecture Search". In: *arXiv e-prints*, arXiv:1910.11858 (Oct. 2019), arXiv:1910.11858. arXiv: `1910.11858 [cs.LG]`.

[400] Hanxiao Liu, Karen Simonyan, and Yiming Yang. "DARTS: Differentiable Architecture Search". In: *arXiv e-prints*, arXiv:1806.09055 (June 2018), arXiv:1806.09055. arXiv: `1806.09055 [cs.LG]`.

[401] Zhongkai Shangguan et al. "Neural Process for Black-Box Model Optimization Under Bayesian Framework". In: *arXiv e-prints*, arXiv:2104.02487 (Apr. 2021), arXiv:2104.02487. arXiv: `2104.02487 [cs.LG]`.

[402] F. Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[403] Donald Jones, Matthias Schonlau, and William Welch. "Efficient Global Optimization of Expensive Black-Box Functions". In: *Journal of Global Optimization* 13 (Dec. 1998), pp. 455–492. DOI: `10.1023/A:1008306431147`.

[404] Niranjan Srinivas et al. "Information-Theoretic Regret Bounds for Gaussian Process Optimization in the Bandit Setting". In: *IEEE Transactions on Information Theory* 58.5 (2012), pp. 3250–3265. DOI: `10.1109/TIT.2011.2182033`.

[405] William R Thompson. "On the Likelihood That One Unknown Probability Exceeds Another in View of the Evidence of Two Samples". In: *Biometrika* 25.3-4 (Dec. 1933), pp. 285–294. ISSN: 0006-3444. DOI: `10.1093/biomet/25.3-4.285`. URL: `https://doi.org/10.1093/biomet/25.3-4.285`.

[406] Daniel Russo et al. "A Tutorial on Thompson Sampling". In: *arXiv e-prints*, arXiv:1707.02038 (July 2017), arXiv:1707.02038. arXiv: `1707.02038 [cs.LG]`.

[407] Qun Meng, Songhao Wang, and Szu Hui Ng. "Combined Global and Local Search for Optimization with Gaussian Process Models". In: *arXiv e-prints*, arXiv:2107.03217 (July 2021), arXiv:2107.03217. arXiv: `2107.03217 [stat.ML]`.

[408] David Eriksson and Matthias Poloczek. "Scalable Constrained Bayesian Optimization". In: *arXiv e-prints*, arXiv:2002.08526 (Feb. 2020), arXiv:2002.08526. arXiv: `2002.08526 [cs.LG]`.

[409] David Eriksson and Martin Jankowiak. "High-Dimensional Bayesian Optimization with Sparse Axis-Aligned Subspaces". In: *arXiv e-prints*, arXiv:2103.00349 (Feb. 2021), arXiv:2103.00349. arXiv: `2103.00349 [cs.LG]`.

[410] Samuel Daulton et al. "Multi-Objective Bayesian Optimization over High-Dimensional Search Spaces". In: *arXiv e-prints*, arXiv:2109.10964 (Sept. 2021), arXiv:2109.10964. arXiv: `2109.10964 [cs.LG]`.

[411] David Wales and Jonathan Doye. "Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters Containing up to 110 Atoms". In: *arXiv e-prints*, cond-mat/9803344 (Mar. 1998), cond–mat/9803344. arXiv: `cond-mat/9803344 [cond-mat]`.

[412] Estefanía Garijo del Río, Jens Jørgen Mortensen, and Karsten Wedel Jacobsen. "Local Bayesian optimizer for atomic structures". In: Phys. Rev. B 100.10, 104103 (Sept. 2019), p. 104103. DOI: `10.1103/PhysRevB.100.104103`. arXiv: `1808.08588 [physics.comp-ph]`.

[413] Jonas Mockus. "Bayesian approach to global optimization: theory and applications". In: 37 (2012).

[414] Marko Järvenpää et al. "Efficient acquisition rules for model-based approximate Bayesian computation". In: *arXiv e-prints*, arXiv:1704.00520 (Apr. 2017), arXiv:1704.00520. arXiv: `1704.00520 [stat.ML]`.

[415] Michael U. Gutmann, Jukka Cor, and er. "Bayesian Optimization for Likelihood-Free Inference of Simulator-Based Statistical Models". In: *Journal of Machine Learning Research* 17.125 (2016), pp. 1–47. URL: `http://jmlr.org/papers/v17/15-017.html`.

[416] Florent Leclercq. "Bayesian optimization for likelihood-free cosmological inference". In: Phys. Rev. D 98.6, 063511 (Sept. 2018), p. 063511. DOI: `10.1103/PhysRevD.98.063511`. arXiv: `1805.07152 [astro-ph.CO]`.

[417] Javier González, Michael Osborne, and Neil D. Lawrence. "GLASSES: Relieving The Myopia Of Bayesian Optimisation". In: *arXiv e-prints*, arXiv:1510.06299 (Oct. 2015), arXiv:1510.06299. arXiv: `1510.06299 [stat.ML]`.

[418] Xubo Yue and Raed AL Kontar. "Why Non-myopic Bayesian Optimization is Promising and How Far Should We Look-ahead? A Study via Rollout". In: Proceedings of Machine Learning Research 108 (26–28 Aug 2020). Ed. by Silvia Chiappa and Roberto Calandra, pp. 2808–2818. URL: https://proceedings.mlr.press/v108/yue20b.html.

[419] Jian Wu and P. Frazier. "Practical Two-Step Lookahead Bayesian Optimization". In: *NeurIPS*. 2019.

[420] Julien Villemonteix, Emmanuel Vazquez, and Eric Walter. "An Informational Approach to the Global Optimization of Expensive-to-Evaluate Functions". In: *J. of Global Optimization* 44.4 (Aug. 2009), pp. 509–534. ISSN: 0925-5001. DOI: 10.1007/s10898-008-9354-2. URL: https://doi.org/10.1007/s10898-008-9354-2.

[421] Philipp Hennig and Christian J. Schuler. "Entropy Search for Information-Efficient Global Optimization". In: *J. Mach. Learn. Res.* 13.null (June 2012), pp. 1809–1837. ISSN: 1532-4435.

[422] Zi Wang and Stefanie Jegelka. "Max-Value Entropy Search for Efficient Bayesian Optimization". In: ICML'17 (2017), pp. 3627–3635.

[423] Louis C. Tiao et al. "BORE: Bayesian Optimization by Density-Ratio Estimation". In: *arXiv e-prints*, arXiv:2102.09009 (Feb. 2021), arXiv:2102.09009. arXiv: 2102.09009 [cs.LG].

[424] George De Ath, Tinkle Chugh, and Alma A. M. Rahat. "MBORE: Multi-objective Bayesian Optimisation by Density-Ratio Estimation". In: *arXiv e-prints*, arXiv:2203.16912 (Mar. 2022), arXiv:2203.16912. arXiv: 2203.16912 [cs.LG].

[425] Jacob R Gardner et al. "GPyTorch: Blackbox Matrix-Matrix Gaussian Process Inference with GPU Acceleration". In: *Advances in Neural Information Processing Systems*. 2018.

[426] Ke Alexander Wang et al. "Exact Gaussian Processes on a Million Data Points". In: *arXiv e-prints*, arXiv:1903.08114 (Mar. 2019), arXiv:1903.08114. arXiv: 1903.08114 [cs.LG].

[427] Nikolaus Hansen et al. *CMA-ES/pycma: r3.2.2*. Version r3.2.2. Mar. 2022. DOI: 10.5281/zenodo.6370326. URL: https://doi.org/10.5281/zenodo.6370326.

[428] Pauli Virtanen et al. "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python". In: *Nature Methods* 17 (2020), pp. 261–272. DOI: 10.1038/s41592-019-0686-2.

[429] Adam Paszke et al. "PyTorch: An Imperative Style, High-Performance Deep Learning Library". In: (2019). Ed. by H. Wallach et al., pp. 8024–8035. URL: http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.

[430]   Diederik P. Kingma and Jimmy Ba. "Adam: A Method for Stochastic Optimization". In: *arXiv e-prints*, arXiv:1412.6980 (Dec. 2014), arXiv:1412.6980. arXiv: `1412.6980` `[cs.LG]`.