

TITLE TODO

LEANDER F. THIELE

A DISSERTATION
PRESENTED TO THE FACULTY
OF PRINCETON UNIVERSITY
IN CANDIDACY FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE
BY THE DEPARTMENT OF
PHYSICS
ADVISER: DAVID N. SPERGEL

FEBRUARY 2024

© Copyright by Leander F. Thiele, 2024.

All rights reserved.

Abstract

Abstract here. Less than 350 words.

Acknowledgements

Acknowledgments here.

DEDICATION TODO

The work described in this dissertation has been published in the following articles
and presented at the following conferences:

chapter 4: *Astrophys. J.* **902**(2), 129 (2020).

chapter 2: *Phys. Rev. D.* **104**(6), 063535 (2021).

chapter 3: *Phys. Rev. D.* **105**(8), 083585 (2022).

chapter 5: *Mach. Learn.: Sci. Technol.* **3**(3), 035002 (2022).

chapter 6: *Phys. Rev. D.* **108**(12), 123526 (2023).

chapter 7: arXiv:2307.07555 submitted to *Astrophys. J.*

American Astronomical Society Meeting, June 2022.

Contents

Abstract	iii
Acknowledgements	iv
List of Tables	xi
List of Figures	xii
1 Introduction	1
1.1 What cosmology is about	1
1.2 The standard model	5
1.3 New frontiers	5
2 Can small-scale baryon inhomogeneities resolve the Hubble tension? An investigation with ACT DR4	7
2.1 Introduction	7
2.2 Setup	11
2.3 Results: posteriors	15
2.4 Results: best-fit models	18
2.5 Conclusions	22
3 Percent-level constraints on baryonic feedback with spectral distortion measurements	25
3.1 Introduction	26
3.2 Simulations	31

3.3	PIXIE experimental model	35
3.4	Interpolating neural networks	35
3.5	Results and Discussion	40
3.6	Conclusions	46
4	Teaching neural networks to generate Fast Sunyaev Zel'dovich Maps	49
4.1	Introduction	49
4.2	Physical background	53
4.3	Methods	55
4.4	Challenge: sparsity	59
4.5	Results	64
4.6	Conclusions	78
5	Predicting the Thermal Sunyaev-Zel'dovich Field using Modular and Equivariant Set-Based Neural Networks	82
5.1	Introduction	83
5.2	Architecture	86
5.3	Data and training	90
5.4	Results and Discussion	94
5.5	Future directions	98
6	Cosmological constraints from HSC Y1 lensing convergence PDF	101
6.1	Introduction	101
6.2	Methods	103
6.3	Results	111
6.4	Conclusions	112
7	Neutrino mass constraint from an Implicit Likelihood Analysis of BOSS voids	114

7.1	Introduction	115
7.2	Simulations	118
7.3	Inference	123
7.4	Results	131
7.5	Conclusions	138
	Bibliography	141
	A Appendix to Chapter 2	184
A.1	Assessment of prior volume effects	184
A.2	Central values and uncertainties	187
	B Appendix to Chapter 3	188
B.1	Electron temperature and the relativistic SZ effect	188
B.2	Cosmology dependence	190
B.3	Reionization contribution	191
B.4	IGM contribution	193
B.5	Two-parameter dependence	194
B.6	Symbolic regression	194
	C Appendix to Chapter 4	198
C.1	Pixelization	198
C.2	Training boxes	199
C.3	Transformations	200
	D Appendix to Chapter 5	201
D.1	Qualitative comparison to CNN	201
	E Appendix to Chapter 6	204
E.1	Information	204

E.2	Data vector choices	205
E.3	Likelihood choices	206
E.4	Tomography	207
F	Appendix to Chapter 7	208
F.1	Halo occupation distribution	208
F.2	Broadening of posteriors	210
F.3	Corner plots	214
F.4	Impact of A_s prior	216
F.5	Another look at the augmentations	216
F.6	Simulation budget	216
F.7	Simulation data	216
F.8	Code	216

List of Tables

2.1	recent H_0 measurements	8
2.2	three-zone model	12
2.3	best-fit models	18
2.4	$\Delta\chi^2$ for best-fit models	22
3.1	simulation results at fixed parameters	34
A.1	posterior characteristics	187

List of Figures

1.1	<i>Planck</i> 2015 power spectra	6
2.1	posteriors with old parameterization and likelihood	14
2.2	posteriors with old parameterization, including ACT likelihood	15
2.3	posteriors with extended parameterization	16
2.4	posteriors with added BAO likelihood	17
2.5	ionization fraction vs. redshift	19
2.6	CMB power spectra	20
3.1	CAMELS predictions compared to PIXIE sensitivity	27
3.2	halo contributions to y distortions	29
3.3	analytical correction factors	37
3.4	dependence on feedback parameters	41
4.1	schematic illustration of dark matter to baryons mapping	52
4.2	U-Net architecture	57
4.3	sparsity of P_e field: power spectrum	60
4.4	sparsity of P_e field: one-point	60
4.5	P_e power spectrum	65
4.6	P_e projected power spectrum	66
4.7	P_e cross-correlation	66
4.8	P_e PDF	66

4.9	P_e bispectrum	66
4.10	individual halos	69
4.11	n_e power spectrum	70
4.12	n_e cross-correlation	70
4.13	n_e PDF	70
4.14	p_e power spectrum	73
4.15	p_e cross-correlation	74
4.16	p_e projected power spectrum	74
4.17	p_e PDF	74
4.18	cross-correlations	76
5.1	modular DeepSets architecture	87
5.2	loss compared to benchmark	94
5.3	individual clusters: P_e images	96
5.4	individual clusters: P_e profiles	96
5.5	P_e PDF	98
6.1	weak lensing convergence PDF	104
6.2	likelihood validation	108
6.3	assessment of systematic errors	108
6.4	posteriors	111
7.1	cosmological simulations	119
7.2	data vector	123
7.3	augmentations	129
7.4	$\sum m_\nu$ posteriors	132
7.5	choice of k_{\max}	132
7.6	coverage plot	133
7.7	quadrupole	134

7.8	real vs. mock posteriors	134
7.9	void size	135
7.10	void-only posteriors	136
7.11	Poissonian void size function	137
A.1	changing f_2	185
A.2	profile likelihood	186
B.1	simulated reionization Compton- y	192
B.2	two-parameter dependences	195
B.3	symbolic regression results	197
D.1	cluster images from CNN approach	202
E.1	compression weights	204
E.2	data vector modifications	205
E.3	likelihood modifications	206
E.4	tomography	207
F.1	averaged mock posteriors	210
F.2	broadening of posteriors: schematic	211
F.3	mock posteriors with larger volume	211
F.4	posteriors in full parameter space	213
F.5	posteriors in different parameter spaces	213
F.6	full EFTofLSS posteriors	214
F.7	A_s prior	215
F.8	$\sum m_\nu$ standard deviation	215
F.9	simulation budget	215

Chapter 1

Introduction

1.1 What cosmology is about

Cosmology is a fundamentally observational science: it is technologically impossible to experimentally approximate the physical conditions responsible for the properties of our universe. Therefore, all we can aim for is to explain what this single experiment that is our universe presents to us. Over the past century, we have been remarkably successful in doing so, resulting in the cosmological standard model. So far, no decisive evidence has emerged to falsify the standard model, although there are interesting hints of disagreement between observations and predictions. The standard model's success has the side-effect that cosmologists tend to blur the line between observational fact and theoretical interpretation. In this first section, I shall therefore discuss some properties of our universe that are so well-established that we can regard them as truths, before turning to the currently accepted theoretical interpretation.

1.1.1 The universe is vast

Humans are not equipped to really “understand” the scales of interest in cosmology. Let me attempt to give some intuition for these distances.

With the bare eye we can see nearby stars, the closest of which are at a distance of about 1 parsec (pc). These stars constitute the immediate neighbourhood of our solar system, which is a tiny part of the Milky Way, our home galaxy. The Milky Way is a disk-shaped accumulation of billions of stars; its center is about 10 kpc (ten thousand parsec) from us.

Galaxies such as the Milky Way form by gravitational collapse of matter, most likely within larger accumulations of dark matter. These dark matter clumps are called halos; the largest halos have sizes of about 1 Mpc (one million parsec). The dark matter halos have not always been there but also formed through gravitational collapse. In this process, small peaks in the density attract matter in their vicinity. For the largest halos, this vicinity reaches about 10 Mpc. Therefore, 10 Mpc is the distance matter can move during the universe's evolution¹ The gravitational collapse of matter into the density peaks is a complicated process during which the participating matter loses any resemblance to its initial distribution. For this reason, cosmology is mainly concerned with scales larger than 10 Mpc. There are many exceptions to this rule, some of which will appear in this thesis.

The small peaks out of which dark matter halos and galaxies grow are not randomly distributed. Their distribution shows patterns that were imprinted during a much earlier time. These patterns have a range of sizes, but the largest are about 100 Mpc.

Finally, the part of the universe that is currently observationally accessible to us has a finite size.² This size is about 10 Gpc (ten thousand Mpc).

¹One might wonder about motion due to initial velocities. It turns out that these decay rather quickly and therefore do not lead to large-scale motion of matter.

²Our current understanding is that this bound is fundamental and due to the finite age of the universe. If the universe went through a bounce gravitational waves can in principle be used to probe larger volumes.

1.1.2 The universe is symmetric

In physics, the word symmetry is used in a more general sense than in everyday language. A system is said to have a certain symmetry if its fundamental character does not change when a certain transformation is applied to it. For example, the letter A has one mirror symmetry because we can mirror it along a vertical axis and obtain the same letter, whereas the letter O has two mirror symmetries (we can also mirror it along a horizontal axis). We can already see from these examples that the precise use of the symmetry concept requires more than just saying that something is symmetric (as would be common in everyday use); instead, the physicist needs to specify *which* symmetry is meant.

When discussing the symmetries of our universe, we need to clarify the scales under consideration. We shall be concerned with relatively large scales, larger than about 100 Mpc. The reason is that on scales below 100 Mpc the previously mentioned patterns lead to coherent variations in the matter density. By averaging the matter field on large scales we can discern the fundamental symmetries.

Observationally, it is a quite robust finding that on large scales the universe is homogeneous and isotropic. This means that at every point in space (homogeneity) the universe looks the same in all directions (isotropy). The assumption that these properties hold is called the cosmological principle. In other words, we can shift and rotate space arbitrarily without altering any physical properties.

Note that these symmetries only refer to the space, the time coordinate is not involved. It is possible that in the far past and the far future even more symmetries exist, incorporating the time coordinate.

Homogeneity and isotropy greatly simplify the description of the universe's large-scale geometry, as we shall see later. However, they also call for a physical explanation: why does our universe have such a simple structure? Currently, the leading theory is the inflationary paradigm, according to which the early universe underwent a period of

rapid expansion which necessarily ends with a homogeneous and isotropic observable universe.

1.1.3 The universe is dark

Looking up at the sky at night, we notice that it is mostly dark. This may appear very natural, but actually has some interesting consequences for cosmology. Indeed, if the universe were static and infinitely large, with a homogeneous distribution of stars on large scales, the night sky would be bright. This argument³ suggests that the universe is evolving (the evidence for homogeneity and sufficient size is quite strong). More on this in the following section.

There is a second sense in which the universe is dark. The matter we can observe with our eyes and telescopes through electromagnetic radiation (light) appears to only constitute about 5 % of the universe's energy content. There is another small contribution from photons (the carriers of light) and a moderate contribution from neutrinos (very light, weakly interacting particles). The main contributions seem to come from dark energy and dark matter, two forms of energy we know very little about.

Dark matter makes itself felt only through its gravitational effects; indeed, it is difficult to explain the existence of galaxies as we observe them today were it not for the extra assistance in gravitational collapse provided by the dark matter.

Dark energy is even more mysterious. Our current theory requires it to explain the recently accelerating expansion of the universe (which is observationally well-established). However, natural explanations for its character make wrong predictions about how much dark energy there should be. Currently, it is possible to explain the data with Einstein's cosmological constant, to be introduced later.

³called Olbers's paradox. In accordance with Stigler's law, Olbers was not the one to first notice it.

1.1.4 The universe is evolving

1.1.5 The universe is complex

1.1.6 The cosmological program

1.2 The standard model

1.2.1 Einstein gravity

1.2.2 Background geometry

1.2.3 Thermal history

1.2.4 Perturbations

1.2.5 Halo model

1.3 New frontiers

1.3.1 Hubble tension

1.3.2 Baryons

1.3.3 Non-Gaussian statistics

1.3.4 Machine learning

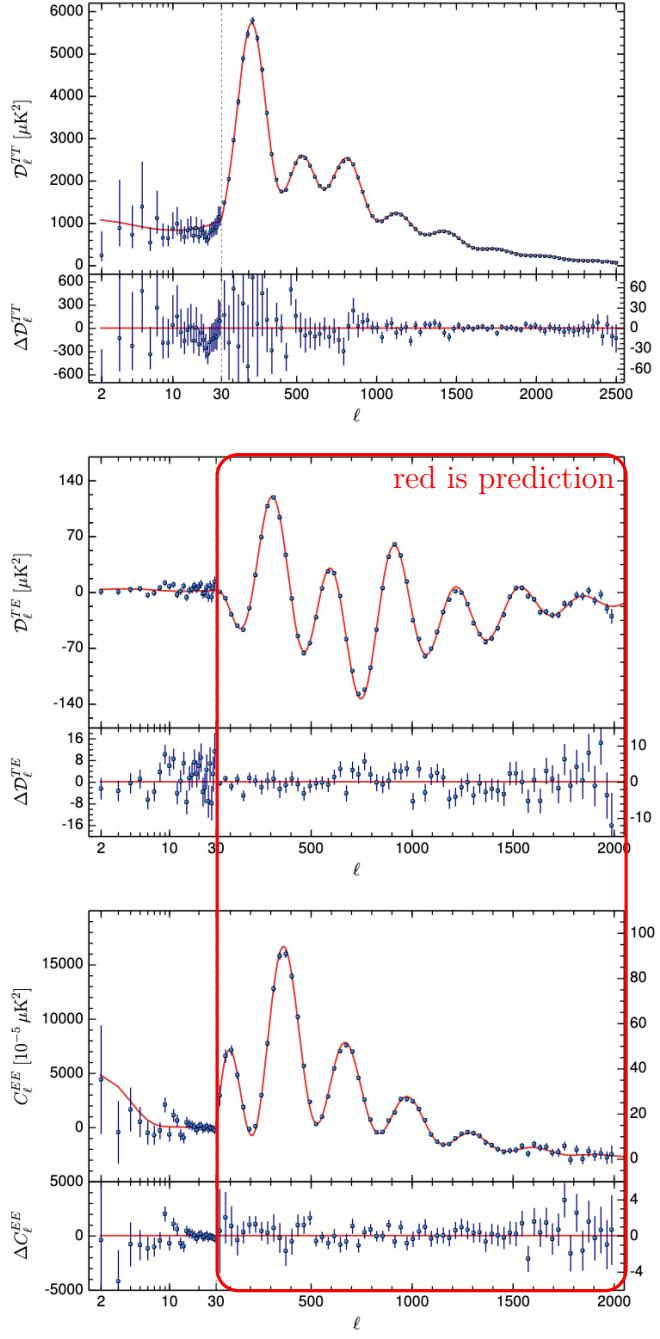


Figure 1.1: Adapted from Ref. [1].

Chapter 2

Can small-scale baryon inhomogeneities resolve the Hubble tension? An investigation with ACT DR4

Abstract. Small-scale inhomogeneities in the baryon density around recombination have been proposed as a solution to the tension between local and global determinations of the Hubble constant. These baryon clumping models make distinct predictions for the cosmic microwave background anisotropy power spectra on small angular scales. We use recent data from the Atacama Cosmology Telescope to test these predictions. No evidence for baryon clumping is found, assuming a range of parameterizations for time-independent baryon density probability distribution functions. The inferred Hubble constant remains in significant tension with the SH0ES measurement.¹

2.1 Introduction

One cornerstone of the Λ CDM standard model of cosmology is the Hubble constant H_0 . We may define it in two ways. In a model-agnostic way, H_0 measures the rate at

¹based on Ref. [2], joint work with Y. Guan, J. C. Hill, A. Kosowsky, and D. N. Spergel.

which space is expanding at the present day. This definition enables us to determine H_0 locally by measuring the apparent speed of recession of galaxies in the Hubble flow. The other definition depends on the assumed cosmological model: it interprets c/H_0 as a measurement of the angular diameter distance to a given redshift, modulo the values of the background density parameters in the Friedmann equation. Measurements of H_0 according to this definition can be performed from the cosmic microwave background (CMB) power spectra or the late-time matter distribution. For both measurements, the baryonic acoustic oscillation (BAO) feature sets a characteristic length scale.

In the correct cosmological model the two definitions of the Hubble constant should be consistent, i.e., barring unknown systematics low- and high-redshift measurements should give consistent results. This, however, appears not to be the case, as is illustrated in Tab. 2.1. Roughly, a $> 4\sigma$ discrepancy between the low- and high-redshift inferences is observed.

Currently, the observed discrepancy is the most severe anomaly in the Λ CDM standard model, and hence the most promising avenue for the discovery of new physics in cosmology. Thus, it has spurred a wide range of attempts at an explanation (so many, in fact, that we now have dedicated reviews to keep track of them, e.g., [12]). Pre-recombination modifications include the Early Dark Energy models [13–18], which however are strongly constrained by large-scale structure data [19–21]. Additional light

	dataset	H_0 [km/s/Mpc]	type
local	SH0ES [3]	73.2 ± 1.3	SNIa+Cepheids
	CSP [4]	$69.6 \pm 0.8 \pm 1.7$	DEB + TRGB
	MCP [5]	73.9 ± 3.0	VLBI geometric
	H0LiCOW [6]	$73.3^{+1.7}_{-1.8}$	lensing time delay
	TDCOSMO [7]	$67.4^{+4.1}_{-3.2}$	lensing time delay
global	<i>Planck</i> [8]	67.4 ± 0.5	CMB
	ACT [9]	67.9 ± 1.5	CMB
	ACT+WMAP [9, 10]	67.6 ± 1.1	CMB
	BOSS [11]	68.6 ± 1.1	BAO

Table 2.1: Selection of recent H_0 measurements. See Sec. 2.2 for comment on the H0LiCOW measurement.

states may also be able to accommodate higher H_0 , as long as their dynamics is non-trivial [22–24] (but see [25, 26]). Post-recombination modifications such as dynamical dark energy and modified gravity [27–38] or a fifth force [39] approach the tension from the opposite point of view, being not always physically well-motivated however, and often encountering problems with low-redshift expansion history data [40].

One attempted resolution of the H_0 -tension involves modifications of Λ CDM around the time of recombination that decrease the sound horizon r_s . The peaks and troughs in the CMB power spectra are located at wave numbers scaling as $(H_0 r_s)^{-1}$. Thus, a decrease in the sound horizon allows for higher values of H_0 as inferred from the CMB. It should be noted that this picture is somewhat simplistic, since only changing r_s introduces other tensions [41]. The model considered in this work, however, has other secondary effects on the CMB power spectra that may be able to circumvent this constraint.

The mechanism to achieve the change in r_s that we will focus on in this paper is the introduction of $O(1)$ inhomogeneities in the baryon density around the time of recombination. These inhomogeneities are required to be on comoving scales well below a Mpc so that they would not be directly visible in the currently available CMB data. Even in simple models of recombination physics it is clear that such small-scale baryon inhomogeneities will shift the surface of last scattering to higher redshift, thus decreasing the sound horizon. In fact, Peebles’s model [42] contains a term $\dot{n}_e \propto -n_e^2$, which upon averaging gives faster recombination in inhomogeneous plasma on account of $\langle n_e^2 \rangle > \langle n_e \rangle^2$ [43].

One attractive proposal to achieve such a scenario is the introduction of primordial magnetic fields (PMFs) [44, 43, 45, 46]. This mechanism would have the advantage that it naturally sustains the required amount of small-scale inhomogeneity through a dynamical process, which in many other conceivable models would be entirely erased through Silk damping. Furthermore, theoretically well-motivated PMF configurations

would lead to inhomogeneities of roughly the required amount on kpc scales. The necessary field strengths would be such that present-day large-scale magnetic fields could be explained with limited invocation of galactic dynamo effects. One open question that could be answered through MHD simulations is the effect of an inverse energy cascade from the turbulent small-scale magnetic fields and baryon inhomogeneities towards larger scales. If such an energy flow was appreciable, it would add additional constraints. We emphasize that this work deals with baryon inhomogeneities in an agnostic way and makes no reference to PMFs; however, it seems difficult to generate the required inhomogeneities without invoking PMFs. Furthermore, most other changes to recombination physics that would lead to appreciable shifts in H_0 are difficult to reconcile with atomic physics constraints [47].

While previous work – Ref. [46] – restricted CMB data to the results from the *Planck* satellite [8], this paper will include more recent data from the Atacama Cosmology Telescope (ACT), specifically their data release 4 (DR4) [48, 9]. Although the angular scales covered by *Planck* enable tight constraints on the location of the acoustic peaks (and thus on $H_0 r_s$), they are less constraining with regard to a secondary prediction of the inhomogeneous models, namely the change in amplitude in the Silk damping tail [43]. While faster recombination decreases the time photons have to diffuse (thus suppressing Silk damping), helium recombination is also sped up which decreases the ionization fraction and therefore increases the photon mean free path (thus enhancing Silk damping). Generically, these two competing effects will not completely cancel; hence the inhomogeneous models have a characteristic signature on small angular scales that is not degenerate with the shift in peak positions. For this reason, the inclusion of small-scale ACT data provides valuable additional constraints on the proposed resolution of the H_0 tension. We may go even further and posit that the new ACT data covering scales that were not considered in [46] provide an independent test of the secondary small-scale effects predicted by this model. If the ACT data do

not show evidence of these effects, then the probability of this model resolving the Hubble tension is accordingly lessened, compared to a hypothetical additional data set that covers the same scales as *Planck*.

As the baryon inhomogeneities are on scales 10^3 times below current resolution, their effect is completely described by the baryon density probability distribution function (PDF). While future MHD simulations assuming PMFs could provide constraints on the shape of this PDF, in this work we treat it as physical input. This has the advantage that we can treat the baryon inhomogeneities agnostically with regard to their origin, but we will have to make certain assumptions on the shape of the PDF in order to retain a predictive model. Future work will hopefully be able to give physically better motivated parameterizations of the PDF; however, we believe that the range of PDFs considered in this work provides strong motivation to believe that our results are in fact generic.

The rest of this paper is structured as follows. In Sec. 2.2, we review the three-zone model for the baryon density PDF from [46] that we adopt in this work. Furthermore, we explain our choice of likelihoods and summarize the methods used. Sec. 2.3 contains the main results of this work, namely posteriors on H_0 and various parameters of the three-zone model for a variety of likelihoods. In Sec. 2.4 we add local H_0 measurements to the CMB likelihoods in order to fully excite the baryon clumping mode and derive some intuition from the resulting best-fit models. We conclude in Sec. 2.5.

2.2 Setup

Following [46], we parameterize the baryon density PDF as a three-zone model, where each zone is described by its density $\Delta_i = n_b^i / \langle n_b \rangle$ and its volume fraction f_i ($i = 1 \dots 3$). We can think of the comoving coordinates being split into small regions, each of which realizes one of the three zones.

model	b	f_2	Δ_1	Δ_2
Λ CDM	0	1	0	1
M1	free	1/3	0.1	1
M2	free	1/3	0.3	1
$f_2 = 1/3$	free	1/3	free	1
$f_2 = 1/2$	free	1/2	free	1
$f_2 = 2/3$	free	2/3	free	1

Table 2.2: Versions of the three-zone model used in this work. M1 and M2 are identical to the ones from [46], while the other three have one more degree of freedom.

We introduce the ‘clumping factor’ b as

$$1 + b = \sum_i f_i \Delta_i^2 \quad (2.1)$$

and remind the reader of the constraints $\sum_i f_i = 1$ and $\sum_i f_i \Delta_i = 1$. The baryon clumping models are defined by four parameters, because there are three zones, each with density and volume fraction, and two constraints between these quantities. For physical convenience, we choose the 4-parameter model space to be spanned by f_2 , Δ_1 , Δ_2 , and the clumping factor b defined in Eq. (2.1).

In this paper, for ease of calculation and comparison with [46], we consider several subsets of these models, with either one or two of the four parameters being free and the others fixed to specific values, as listed in Tab. 2.2. These likely bracket all of the reasonable physical cases, but a more extensive analysis could consider 4-parameter models instead.

While models M1 and M2 were chosen to facilitate direct comparison with [46], the more extended models ($f_2 = \dots$) are constructed so as to cover a relatively wide range of possible baryon density PDFs while also being convenient computationally². They also constrain the variations of the baryon density to $O(1)$ excursions, consistent with the prediction by the PMF model.

We use two combinations of Boltzmann codes, recombination models, and Monte Carlo samplers to independently arrive at our results, namely **CAMB** [49] + **RECFAST** [50–

²Not all parameterizations are equally good, some will only allow narrow windows in the free parameters in which the constraint system has solutions at all, which complicates the sampling and interpretation.

[52\]](#) + [Cobaya](#) [53, 54] and [CLASS](#) [55, 56] + [HyRec](#) [57] + [MontePython](#) [58, 59]³. As previous work on our topic had relied on [RECFEST](#) which was calibrated for models close to Λ CDM, and certainly not for $O(1)$ deviations from the average baryon density, an independent check with the more sophisticated [HyRec](#) code is reassuring. All figures in this work were produced with the second toolchain. We run [CLASS](#) with the provided `cl_permil.e.pre` file instead of the default precision settings, and second Ref. [60] in recommending this as good practice, given the high precision of current and upcoming CMB data⁴. Our convergence criterion is for the Gelman-Rubin R_α to achieve $\max_\alpha(R_\alpha - 1) < 0.05$ (α labels the parameters), and we generally achieve $R_{H_0} - 1 \lesssim 0.02$.

In both codes, we separately compute the recombination history $x_e^i(z)$ in each of the three zones and then perform the average to arrive at the large-scale ionization fraction

$$\langle x_e(z) \rangle = \sum_i f_i \Delta_i x_e^i(z). \quad (2.2)$$

We use the following likelihoods:

- *Planck* 2018: high- ℓ TTTEEE (`lite`), low- ℓ TT, low- ℓ EE [61];
- ACT DR4: high- ℓ TT, EE, TE [9, 48];
- SMH: SH0ES [62], MCP [5], H0LiCOW [6];
- BAO: SDSS DR7 [63] & BOSS DR9 [64], 6dFGS [65].

We found that working with the extended *Planck* likelihood (as opposed to the `lite` version) did not give appreciable changes in the posteriors in our parameter space.

The ACT likelihood is chosen such that covariance with *Planck* is negligible, with $\ell < 1800$ discarded in TT but no cuts in TE and EE [9].

³the latter with the flag `MODEL=FULL` instead of the default `RECFEST` that ships with [CLASS](#).

⁴We performed a test with the default precision file on the $f_2 = 1/2$ model. In that case, there was a negligible $\sim 0.1\sigma$ bias on H_0 , while the shape of the b -posterior was slightly but visibly altered. Thus, the precision does not seem to be an issue in this work's context, but this is not generally true in extensions of Λ CDM constrained with post-*Planck* experiments.

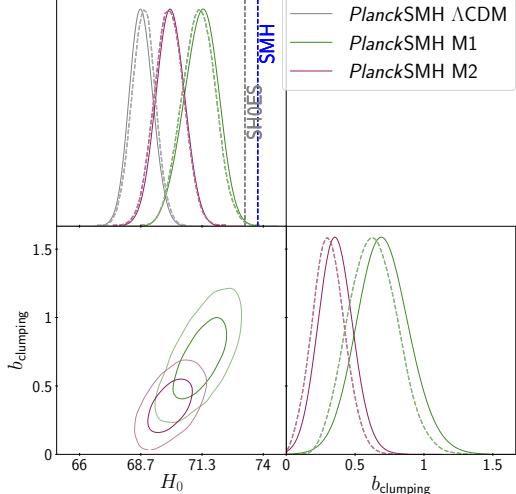


Figure 2.1: Posteriors in the same parameterizations and with the same likelihood as the one used in [46]. The solid lines are our results, while the dashed lines are [46]. Here and in all subsequent posterior plots, the grey shaded region is the 1σ confidence interval from the latest SH0ES measurement [3]. We have also indicated the 1σ confidence interval of the SMH H_0 likelihood used in this plot and in Sec. 2.4.

The H0LiCOW measurement used in the SMH likelihood is not the most recent one; an updated lens modeling found substantially lower H_0 [7]. Similarly, we do not use the most recent SH0ES measurement [3] but rather the one used in [46] from Ref. [62]. The reason for this is that we use the SMH combination of likelihoods only for direct comparison to [46] (equal to their ‘‘H3’’) as well as in contexts in which we simply want to excite the baryon clumping mode to a considerable extent (i.e., not in posteriors).

We assume flat priors on ω_b , ω_{CDM} , n_s , A_s , H_0 , τ_{reio} , A_{Planck} and the hard prior $0.9 < y_p < 1.1$ for the only ACT nuisance parameter (polarization efficiency). When they are used, the three-zone model parameters have the wide priors $0 < b < 10$ and $0 < \Delta_1 < 0.7$ (the upper bound on Δ_1 is important as it removes the spurious second posterior mode that would otherwise be allowed by inverting the first and third zone). We assume a fixed neutrino mass sum of 0.06 eV (this helps us focus on the degeneracy between H_0 and baryon clumping, rather than having to worry about the $H_0 - m_\nu$ degeneracy in addition).

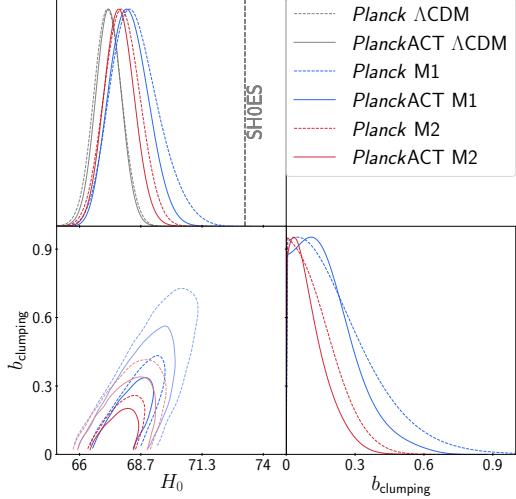


Figure 2.2: Posteriors in the same parameterizations as in [46], comparing *Planck* alone with ACT added. Here and in the following posteriors, the line color indicates the model while the line style differentiates between different likelihoods.

2.3 Results: posteriors

As a first step, we reproduce the results from [46], staying within their M1/M2 parameterization of the three-zone model and working solely with the *Planck*+SMH likelihood (which we adopt only in order to reproduce previous results and assess the accuracy of RECFAST in the baryon clumping models). The resulting posteriors are shown in Fig. 2.1. We observe that our posteriors on H_0 are almost identical with those found in [46] (which are plotted as dashed lines). There are some minor discrepancies in the posterior on the clumping parameter b , which we believe are most likely explained by differences between HyRec and RECFAST⁵.

In the remainder of this section, we will not include SMH (i.e. SH0ES, MCP, and the old H0LiCOW value) in the likelihoods. As we will see, the CMB(+BAO) likelihoods are not compatible with SMH even after allowing for baryon clumping, which in our view renders a formally combined likelihood problematic for interpretation.

In Fig. 2.2, we compare posteriors in the [46] parameterizations fitted to *Planck* and *Planck*+ACT. For these posteriors, as well as for all the others plotted in this section, central values and $1-\sigma$ uncertainties are listed in Tab. A.1 in Appendix A.2. First,

⁵As mentioned in Sec. 2.2, we confirmed key results with the tool-chain used in [46]. In that case, we find almost perfect agreement with their results.

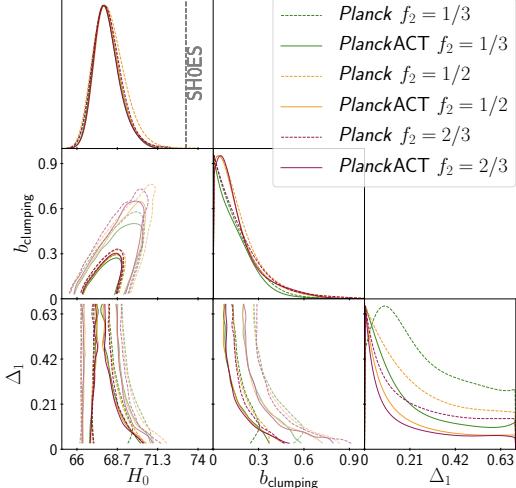


Figure 2.3: Posteriors in the parameterizations with one more degree of freedom, Δ_1 .

we observe that within ΛCDM both likelihoods yield very similar posteriors on H_0 (this is to be expected, as ACT DR4 itself has only about 1/3 the statistical power of *Planck*, and moreover their ΛCDM constraints independently agree well [9]). However, in the models allowing baryon clumping, M1 and M2, the situation is different. While the H_0 posterior's lower end is virtually identical for both likelihoods, the inclusion of ACT significantly reduces the extent of the high- H_0 tail (due to this asymmetry, the posterior plot is more illuminating than the bare $1-\sigma$ uncertainties). The effect is more pronounced for model M1 which yields a less severe tension with SH0ES than M2 does. Similarly, adding ACT reduces the 95 % upper limits on the clumping parameter b (c.f. Tab. A.1). Although M1 and M2 consistently lead to slightly higher central values of H_0 , the shift is not nearly enough to get in reasonable agreement with the SH0ES measurement. In summary, at least in the [46] parameterizations M1/M2, adding ACT to *Planck* increases the tension with SH0ES within the baryon clumping models.

We proceed by adding an additional degree of freedom to the baryon density PDF, treating Δ_1 as a free parameter. The models considered here may be considered as spanning the entire range of which M1 and M2 sample only two discrete points, while also sampling f_2 (which was identical for M1 and M2) at three different points. Posteriors for the *Planck* and *Planck*+ACT likelihoods are shown in Fig. 2.3. We

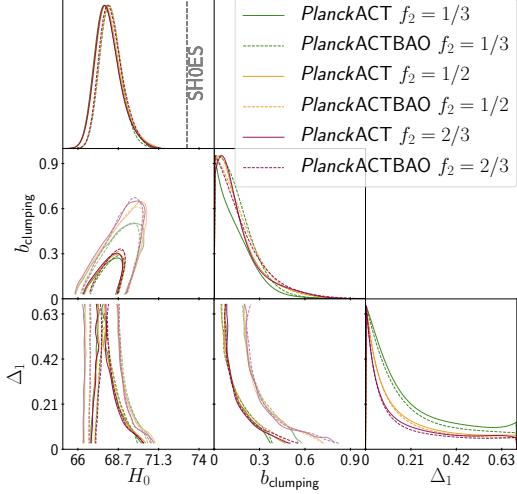


Figure 2.4: Posteriors in the extended parameterization, illustrating the effect of adding a BAO likelihood.

observe extremely similar posteriors in H_0 and b for the three flavors of this extended parameterization. Again, no detection of the clumping effect (i.e., non-zero b) is made. Similarly to the posteriors shown in Fig. 2.2, the addition of ACT has the most pronounced effect in the high- H_0 tails, where it visibly reduces the leeway the baryon clumping models have in getting into somewhat decent agreement with SH0ES.

The extra degree of freedom introduced in these models, Δ_1 , is virtually unconstrained (for this reason, we do not include it in Tab. A.1). This may be a consequence of a poor choice of parameterization, however, it seems more likely that we are already close to covering the entire range of CMB spectra compatible with the data and allowed by the three-zone model. This view is supported by the fact that all three values of f_2 lead to extremely similar posteriors. If this is indeed the case, the H_0 posteriors shown in Fig. 2.3 should be similar to those that could be obtained with any reasonable baryon density PDF. As we discuss in the conclusions, time dependence may prove to be an additional mode that could alter this picture.

We discuss a possible prior volume effect in the more extended $f_2 = \dots$ models in Appendix A.1.

As a final result of this section, we show in Fig. 2.4 how adding $z \lesssim 1$ BAO data affects the posteriors in the extended models. As for H_0 the BAO likelihood has

likelihood model	<i>Planck+ACT</i>	<i>Planck+ACT+SMH</i>		
	ΛCDM	ΛCDM	M1	$f_2 = 1/2$
H_0	67.200	68.703	70.342	72.178
$100\omega_b$	2.2316	2.2561	2.2511	2.2520
ω_{cdm}	0.1204	0.1171	0.1203	0.1246
n_s	0.9660	0.9742	0.9652	0.9523
$10^9 A_s$	2.1086	2.1282	2.0940	2.0617
τ_{reio}	0.0522	0.0687	0.0533	0.0484
b	—	—	0.4071	1.1822
f_1	—	—	0.2865	0.3524
Δ_1	—	—	0.1000	0.0048
f_2	—	—	0.3333	0.5000
Δ_2	—	—	1.0000	1.0000
f_3	—	—	0.3801	0.1476
Δ_3	—	—	1.6784	3.3757
A_{Planck}	1.0043	1.0042	1.0030	1.0018
y_p	1.0012	1.0016	1.0026	1.0016

Table 2.3: Best-fit models used in Sec. 2.4. The first group of parameters are ΛCDM , the second group describes the three-zone model (not all of them are independent), and the third group are the nuisance parameters for our likelihoods.

comparable statistical power to the CMB data, the observation that the low- H_0 tail is reduced by the addition of BAO is no surprise. On the other hand, the high- H_0 tail is almost invariant under adding BAO, which indicates that the CMB data alone are almost entirely responsible for constraining the baryon clumping modes of the model. In fact, the posterior on b is slightly widened once BAO is included. Thus, we conclude that the BAO likelihood is not particularly useful in constraining the baryon clumping models, at least as long as the CMB-only likelihood does not allow appreciable increases in H_0 .

2.4 Results: best-fit models

In order to gain some further intuition, in this section we attempt to fully excite the baryon clumping mode and explore its consequences for the CMB spectra. In order to do this, we now add the previously discussed SMH likelihood to *Planck+ACT*. For

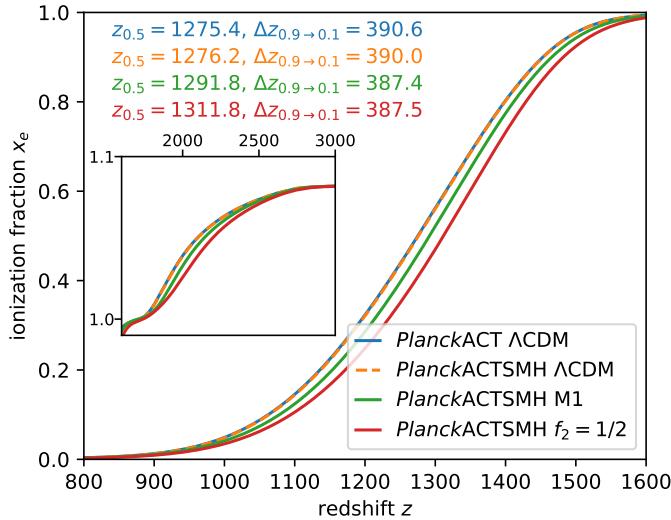


Figure 2.5: The ionization fraction as a function of redshift for the best-fit models listed in Tab. 2.3.

this combined likelihood, we compute three best-fit models, listed in Tab. 2.3 (the model in the first column, *Planck*+ACT Λ CDM, was not fitted to SMH and serves as the reference point). The models including baryon clumping, M1 and $f_1 = 1/2$, were chosen because they are the ones that had the highest central values for H_0 in the posteriors discussed in the previous section. We emphasize that it can be difficult to infer general trends from individual best-fit models, so the results in this section should be taken as starting points to develop intuition. The main results of this work have been presented in the previous section.

For each of the best-fit models, we compute the average ionization fraction $x_e(z)$, shown in Fig. 2.5. In the figure, we also list $z_{0.5}$ and $\Delta z_{0.9 \rightarrow 0.1}$ as proxies for the location and width of the last scattering surface. First, we observe that without inclusion of baryon inhomogeneities the recombination history is essentially fixed. As expected, once the additional degrees of freedom in the baryon density PDF are introduced, recombination is pushed to higher redshift. If we write $A \delta \log \Delta z_{0.9 \rightarrow 0.1} = \delta \log z_{0.5}$ (where δ is with respect to the *Planck*+ACT Λ CDM reference model), we find $A = -1.60$ and $A = -3.53$ for M1 and $f_2 = 1/2$ respectively. This is, to order of magnitude, consistent with $A = -2$, the expectation from the argument that the duration of recombination should scale as $v_{\text{rms}}^{-1} \propto T^{-1/2} \propto z^{-1/2}$. However, it

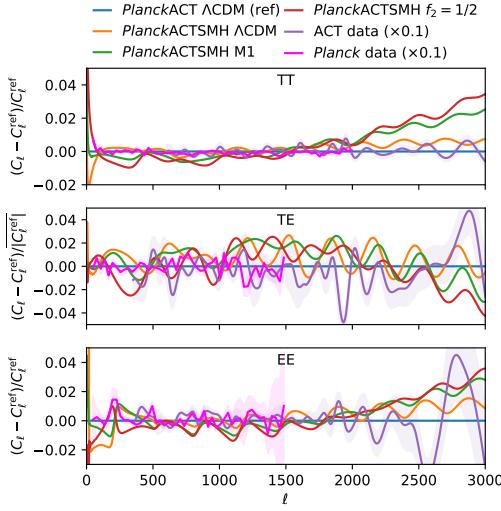


Figure 2.6: Fractional changes in CMB power spectra for the best-fit models listed in Tab. 2.3, with respect to the *Planck*+ACT Λ CDM baseline. In the second panel (TE), because the data have zero crossings, we computed the normalizing factor $|C_\ell^{\text{ref}}|$ as a convolution of the absolute value of the reference TE power spectrum with a Gaussian filter of width $\sigma = 100$ (in ℓ). This gives a useful normalization without artifacts from the roots. Note that for clarity the experimental data curves for ACT and *Planck* have been rescaled by the given factors. While the *Planck* data points are relatively close to what we use in our likelihood (except for binning and ℓ -cuts to improve clarity, in particular for TE and EE), the ACT data points are for illustration purposes only and were generated as explained in the footnote.⁵

is interesting to note that both M1 and $f_2 = 1/2$ yield almost identical values for $\Delta z_{0.9 \rightarrow 0.1}$.

Next, we consider the CMB power spectra arising from the listed best-fit models, shown in Fig. 2.6, where we compare them to *Planck* and ACT data. As already alluded to in Sec. 2.1, the models with baryon inhomogeneities show significant deviations from Λ CDM in the high ℓ part of the spectra, particularly for TT and EE. These deviations are partly caused by shifts in other parameters, as is already evident from the orange curve depicting Λ CDM fit to *Planck*+ACT+SMH. However, in simple experiments where we put baryon inhomogeneities of magnitude comparable to the

⁵The ACT DR4 data only contain band powers, which are related to the underlying CMB power spectra by $C_b = W_{b\ell}C_\ell$, where $W_{b\ell}$ is an $m \times n$ matrix with $m < n$. Thus, it is not possible to uniquely infer the C_ℓ from the band powers. In order to generate power spectra that can be directly compared to *Planck* and the theory models, we first find some solution C_ℓ^0 of the above equation (which is, of course, completely unphysical). Then, writing $K_{\ell\alpha}$ for the $n \times (n-m)$ matrix whose columns are a basis for the kernel of $W_{b\ell}$, we can generate additional solutions $C_\ell(a_\alpha) = C_\ell^0 + K_{\ell\alpha}a_\alpha$, where a_α is an arbitrary vector. We now find the a_α that minimizes $\|\Lambda_{\ell'\ell}C_\ell(a_\alpha)\|$, where $\Lambda_{\ell'\ell}$ implements a discrete second derivative. We perform this procedure for both the ‘deep’ and ‘wide’ parts of the ACT data set, associating approximate error bars with the resulting C_ℓ according to the normalized diagonal elements of the covariance matrix of the C_b . Finally, we take the weighted average of the ‘deep’ and ‘wide’ C_ℓ .

models shown here into otherwise unchanged Λ CDM cosmologies, similar trends occur (the same happens if H_0 is increased so as to neutralize the shift in peak positions). In particular, we always observe an *increase* in TT power at high- ℓ once the baryon inhomogeneities are introduced. Thus, we conclude that the effect of Silk damping is in fact reduced in the inhomogeneous models. This is contrary to previous statements in the literature, which presumably rested on numerical experiments made with much higher amounts of inhomogeneity.

It is interesting to disentangle how the suppression in the $\ell < 1000$ part of the EE spectrum is generated. The ratio of EE to TT for $\ell \gg 1$ can be approximated as [66]

$$\frac{C_\ell^{EE}}{C_\ell^{TT}} = \left(\frac{\eta_* \Delta\eta_*}{r^2(1+R)} \right)^2 B_\ell, \quad (2.3)$$

where η_* and $\Delta\eta_*$ are the location and thickness of the last scattering surface in conformal time respectively, r is its distance from us, $R = 3\rho_b/4\rho_\gamma$, and B_ℓ is a geometric factor independent of cosmology. Comparing, for example, the best-fit models *Planck*+ACT Λ CDM and *Planck*+ACT+SMH $f_2 = 1/2$, we find that the differences in the various quantities entering Eq. (2.3) are all $\sim 2 \dots 4\%$. On aggregate, however, the cancellation gives a $\sim 1\%$ effect.

The ACT data do not seem to show a preference for the general trends in the spectra introduced by the baryon inhomogeneities.

Finally, we present the changes in χ^2 relative to the *Planck*+ACT Λ CDM baseline for the other three best-fit models, listed in Tab. 2.4. The model M1 increases χ^2 by about equal amounts in *Planck* and ACT, while the model $f_2 = 1/2$ with one extra degree of freedom in the baryon density PDF is comparably more disfavoured by *Planck* than by ACT. ACT TT actually appears to favour the baryon clumping models, which holds true for both the entire ℓ -range covered by ACT and $\ell > 1800$, the part used in our likelihoods. On the other hand, ACT TE and, to a lesser extent, EE,

likelihood model	<i>Planck+ACT+SMH</i>		
	ΛCDM	M1	$f_2 = 1/2$
<i>Planck</i> high- ℓ TTTEEE	5.16	3.35	1.82
<i>Planck</i> low- ℓ TT	-1.24	0.31	2.86
<i>Planck</i> low- ℓ EE	1.53	0.02	-0.18
ACT TTTEEE	1.12	2.13	-0.20
ACT TT	-1.82	-3.57	-3.72
ACT TE	3.07	5.52	3.29
ACT EE	-0.33	1.08	2.17
ACT high- ℓ TT	-1.66	-3.09	-2.77
ACT high- ℓ TT+TEEE	1.29	3.46	1.79
SH0ES	-9.06	-16.39	-21.43
MCP	-1.99	-3.58	-4.66
H0LiCOW	-5.25	-9.29	-11.74
CMB likelihood	6.74	7.14	6.29
SMH likelihood	-16.30	-29.26	-37.83

Table 2.4: Values of $\Delta\chi^2$ for three best-fit models from Tab. 2.3, compared to the *Planck+ACT* ΛCDM reference. The final two lines show the combined *Planck+ACT* and SMH likelihoods used in this work.

disfavour baryon clumping. This may be explained by the inability of those models to keep the last scattering surface's thickness fixed (c.f. Fig. 2.5). Another possibility is simply that the steeper slopes in the polarization spectra translate into higher sensitivity to H_0 . It should be mentioned that there have been hints of inconsistencies between ACT TE and other CMB data [9]. As no mechanism has been proposed that could explain these trends, and the tensions are still relatively mild, we do not see this work as the appropriate place to investigate them further.

2.5 Conclusions

The introduction of inhomogeneities in the baryon density around the time of recombination is an attractive proposal to resolve the Hubble tension. While we do not constrain ourselves to a particular mechanism that is supposed to generate the required small-scale power, PMFs appear to be the most natural candidate.

A key prediction of the inhomogeneous models is, besides the geometric shift in acoustic peak positions, a change in the high- ℓ part of the CMB power spectra due to different amount of Silk damping. We have argued that the addition of small-scale ACT data to the previously considered *Planck* likelihood constitutes a powerful way to test this prediction. The higher statistical power of the thus enlarged likelihood also enables us to extend the parameter space; specifically, we are able to add an additional degree of freedom to the baryon density PDF.

Fitting the baryon clumping models to CMB-only likelihoods, we have shown that the inclusion of ACT narrows the posterior for the clumping parameter b around zero compared to *Planck* alone. Likewise, although ACT does not appreciably alter the low- H_0 tails in the posteriors, it significantly shrinks the high- H_0 tails, indicating that the added small-scale information disfavours the introduction of baryon clumping. The CMB data still leave room for mild baryon clumping and some upward shift in H_0 , thus somewhat relieving the Hubble tension, but values as high as the SH0ES 1- σ lower bound are strongly disfavoured.

Although the absence of theoretical predictions for the baryon density PDF restricts us to relatively simple parameterizations, we have argued that there is evidence that the posteriors on H_0 presented in this work are in fact generic for any time-independent baryon density PDF. An intriguing possibility that could evade the constraints presented would be the introduction of time dependence. Assuming the existence of PMFs, the decrease in magnetic field strength during recombination would lead to a homogenization of the baryon distribution (the relaxation time for kpc-scale baryon clumps is considerably smaller than recombination's duration). It is possible that, in the large-scale average, this would lead to a faster onset of recombination while keeping the thickness of the last scattering surface approximately fixed, perhaps leading to a better fit with the polarization data.

Besides our primary results, we have also shown that while the previously used RECFAST recombination code is probably sufficiently accurate for the modifications considered in this work, there are some systematic differences to the more recent `HyRec` which we recommend should be used going forward. Furthermore, we have shown that, contrary to previous claims, baryon inhomogeneities of the amount necessary to appreciably shift H_0 do not generically lead to enhanced Silk damping.

We have found no evidence that the addition of small-scale ACT data to the likelihood favours the introduction of baryon inhomogeneities around recombination over Λ CDM. Looking forward, the increased small-scale CMB information from future ACT releases, SPT-3G [67], Simons Observatory [68], and CMB-S4 [69] will provide more constraining power. As long as no currently unknown systematics in ACT are discovered, however, smaller error bars will only serve to exclude the class of time-independent three-zone models considered in this paper with higher significance. Parallel to this improvement in data quality, theoretical advancements in constructing viable models of the baryon density PDF, for example through MHD simulations, will be essential in increasing the predictivity of the theory.

Modifications to `CLASS` used in this work are publicly available [here](#).

Acknowledgements. We thank Karsten Jedamzik and Levon Pogosian for useful discussions during the final stages of this work. JCH and DNS thank the Simons Foundation for support. LT thanks Zack Li for help with the ACT DR4 likelihood.

Chapter 3

Percent-level constraints on baryonic feedback with spectral distortion measurements

Abstract. High-significance measurements of the monopole thermal Sunyaev-Zel'dovich CMB spectral distortions have the potential to tightly constrain poorly understood baryonic feedback processes. The sky-averaged Compton- y distortion and its relativistic correction are measures of the total thermal energy in electrons in the observable universe and their mean temperature.

We use the CAMELS suite of hydrodynamic simulations to explore possible constraints on parameters describing the subgrid implementation of feedback from active galactic nuclei and supernovae, assuming a PIXIE-like measurement.

The small $25 h^{-1}\text{Mpc}$ CAMELS boxes present challenges due to significant sample variance. We utilize machine learning to construct interpolators through the noisy simulation data. Using the halo model, we translate the simulation halo mass functions into correction factors to reduce sample variance where required.

Our results depend on the subgrid model. In the case of IllustrisTNG, we find that the best-determined parameter combination can be measured to $\simeq 2\%$ and corresponds to a product of AGN and SN feedback. In the case of SIMBA, the tightest constraint is $\simeq 0.2\%$ on a ratio between AGN and SN feedback. A second orthogonal parameter combination can be measured to $\simeq 8\%$. Our results demonstrate the significant constraining power a measurement of the late-time spectral distortion monopoles would have for baryonic feedback models.¹

¹based on Ref. [70], joint work with D. Wadekar, J. C. Hill, N. Battaglia, J. Chluba, F. Villaescusa-Navarro, L. Hernquist, M. Vogelsberger, D. Anglés-Alcázar, and F. Marinacci.

3.1 Introduction

A robust prediction of the standard model of cosmology is the presence of subtle deviations from the blackbody spectrum in the cosmic microwave background (CMB) [e.g., 71, 72]. Such monopole spectral distortions can be caused by a variety of physical processes in both the early- and late-time Universe. Among these various distortion signals, most easily accessible to near-future experiments will be the y -type distortion.² The y distortion arises from inverse Compton scattering of CMB photons with hot electrons, the thermal Sunyaev-Zel'dovich (tSZ) effect [73, 74]. The y signal is mostly sourced by massive, collapsed structures at $z \lesssim 2$. Thus, the tSZ effect informs us about two aspects: First, the thermodynamic properties of the hot electron gas in massive halos; these are very sensitive to astrophysical small-scale processes. Second, the abundance of such halos, which translates into a constraint on the amount and clustering of matter in the Universe. This makes SZ cluster measurements a unique tool for cosmology and astrophysics [75, 76].

To illustrate this point, consider Fig. 3.1. There, we plot a range of theory predictions (blue and magenta points) from the CAMELS suite of hydrodynamic simulations [77]. The axes are the two distortion monopole observables considered in this work, further discussed below. Between the plotted simulations the astrophysical sub-grid model differs, translating into a large spread of theory predictions.³ For comparison, we also show a covariance matrix for a near-future monopole distortion measurement (c.f. Sec. 3.3). The forecast measurement errors are tiny compared to the current theoretical uncertainty, which means that a near-future measurement would provide a substantial gain in information on astrophysics. This basic observation forms the underpinning of the calculations performed in the following: forecast constraints

²For conciseness, we will use the term “ y distortion” or similar to mean both the non-relativistic and the relativistic effects.

³For IllustrisTNG, the spread is dominated by sample variance (as can be seen by comparing the LH and 1P points), a point that we will return to later.

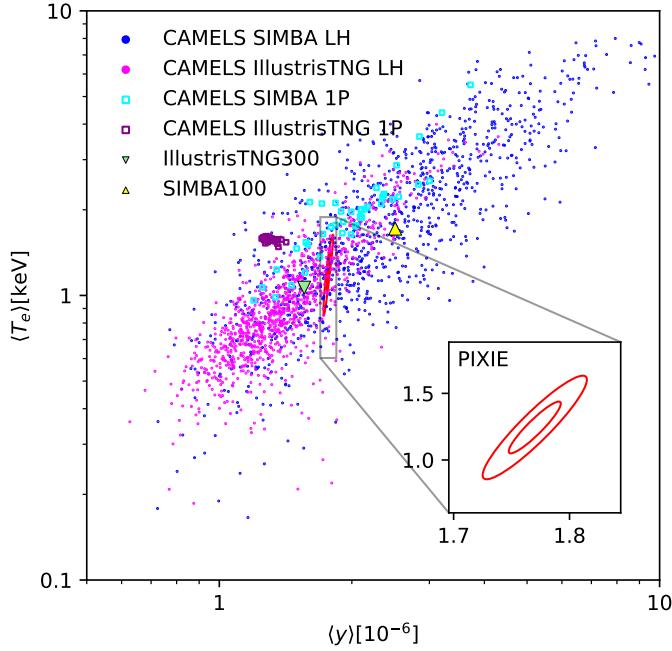


Figure 3.1: Comparison between the range of CAMELS simulations with the forecast covariance matrix for the PIXIE experiment (the PIXIE ellipses are centered at the fiducial model assumed in Ref. [78]: $\langle y \rangle_{\text{fid}} = 1.77 \times 10^{-6}$, $\langle T_e \rangle_{\text{fid}} = 1.24 \text{ keV}$; they correspond to 68 and 95 % CL). From CAMELS, we plot both the LH and the 1P set. The LH data points are contaminated by sample variance in addition to varying subgrid parameters, while the 1P points have the same initial conditions and only differ by their subgrid parameters. We have also indicated the values measured from the large boxes (triangles), as discussed in Sec. 3.2.2. All simulation data have been rescaled to the fiducial CAMELS cosmology but each data point corresponds to a different subgrid model. It is worth noting that the 1P initial conditions appear to be slightly atypical, with generally larger $\langle T_e \rangle$ than expected from the LH set.

on simulation subgrid models from a measurement of the y distortion monopoles, using simulations from the CAMELS suite.

Conventionally, the y -distortion is separated into a non-relativistic and a relativistic [79–83] component, each having a distinct spectral signature that makes it possible to disentangle them observationally.⁴ The non-relativistic contribution is determined by

⁴All these signals can be accurately modeled using `SZpack` [84].

a line-of-sight integral over electron pressure,⁵

$$\langle y \rangle \equiv \langle y(\hat{n}) \rangle_{\hat{n}} = \int \frac{d\hat{n}}{4\pi} \frac{\sigma_T}{m_e} \int P_e(\hat{n}, l) dl, \quad (3.1)$$

where σ_T denotes the Thomson cross section, m_e is electron mass, \hat{n} is the line of sight, and l is physical length along the line of sight. To leading order, the relativistic component is proportional to the y -weighted mean electron temperature [85]

$$\langle T_e \rangle \equiv \langle T_e(\hat{n}) \rangle_{\hat{n}} = \langle y \rangle^{-1} \int \frac{d\hat{n}}{4\pi} \frac{\sigma_T}{m_e} \int [T_e P_e](\hat{n}, l) dl. \quad (3.2)$$

This effective temperature is typically higher than the mass-weighted (or τ -weighted) temperature [86, 87], and can be directly obtained from a moment expansion of the SZ signal (see Appendix B.1 for additional discussion).

In this work we focus on the dominant and theoretically well-established contributions to the distortion signals. In particular, we will neglect the $\simeq 10\%$ contribution of reionization to $\langle y \rangle$, which we discuss in Appendix B.3, as well as signals due to the Milky Way and Local Group (which are estimated to be another one and two orders of magnitude below the reionization signal, respectively). We also neglect other, more exotic sources of y -distortions, for example from primordial magnetic field heating [e.g., 88–90] or decaying particles [e.g., 91–96]. Conversely, if one is interested in using the y -distortions to constrain or detect such processes beyond standard Λ CDM, astrophysical feedback must be very well understood.

There are further relativistic corrections involving higher moments of the electron temperature, as well as the kinetic Sunyaev-Zel'dovich effect sourced by coherent motion; we will neglect these complications and simply treat $\langle y \rangle$ and $\langle T_e \rangle$ as observables, as in Ref. [85].

⁵We set the speed of light and Boltzmann's constant to one.

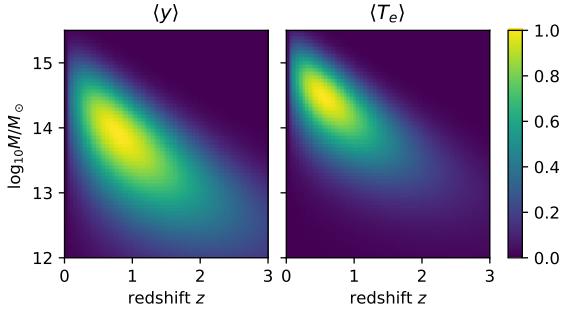


Figure 3.2: Illustration of the halo mass and redshift contributions to the y distortion observables. Plotted is the normalized integrand in Eq. (3.4), using the fitting formulae described in Sec. 3.4.2.

Locally, the tSZ effect is well-established observationally as a CMB temperature change correlating with the locations of clusters. However, the global distortion to the CMB spectrum has not yet been detected. Only an upper limit on the non-relativistic y distortion exists from the COBE FIRAS experiment, which yielded $|\langle y \rangle| < 15 \times 10^{-6}$ (95 % cl) [97], which is about one order of magnitude above the expected Λ CDM signal [85]. We thus anticipate significant detections with future CMB spectroscopy [98–101].

A monopole measurement would be complementary to the existing higher-moment tSZ analyses [e.g., 102–116], since it features very different systematics and would also yield the relativistic component at high significance [78]. Furthermore, in contrast to cluster-stacking and power spectrum [117–119] approaches, the $\langle y \rangle$ measurement is more sensitive to lower-mass objects, as illustrated in Fig. 3.2.

The $\langle y \rangle$ and $\langle T_e \rangle$ signals constitute unique probes of baryonic physics in galaxy clusters and groups. Since $\langle y \rangle$ probes thermal energy, it is subject to the energy conservation equation

$$\underbrace{E_e^{\text{th,tot}}}_{\text{Total}} = \underbrace{E_e^{\text{coll}}}_{\text{Collapse}} + \underbrace{E_e^{\text{inj}}}_{\text{Injected}} - \underbrace{E_e^{\text{cool}}}_{\text{Cooling}} . \quad (3.3)$$

The most uncertain term in the above equation is E_e^{inj} which can largely be attributed to feedback processes from massive stars, supernovae (SNe), and active galactic nuclei (AGN). These processes inject additional energy into the interstellar, intergalactic, and

intracluster media (ISM, IGM, and ICM, respectively). Such feedback processes are standard ingredients in any theoretical models of galaxy formation, both semi-analytic [e.g., 120, 121] and simulation-based models [e.g., 122–127].

The most reliable way to explore how feedback models influence the y -distortions is by analyzing hydrodynamical simulations with qualitatively and quantitatively different subgrid prescriptions. Such an approach is complicated by the fact that, owing to their bias towards rare high-density peaks, the distortion signals are heavily influenced by sample variance. Furthermore, the parameter space of subgrid models is vast and poorly explored, meaning that ideally we would need many large-volume hydrodynamical simulations, which is currently not feasible. We will demonstrate later that these problems can be overcome by utilizing machine learning methods as well as analytical corrections using the halo model.

Since the y -distortions are predominantly sourced by galaxy groups and clusters [85], an analytical description based on the halo model [128–130] is a natural first approximation. In the halo model formalism, we assume spherically symmetric halos described only by mass M and redshift z , yielding

$$\langle y \rangle_{\text{hm}} = \frac{\sigma_8}{m_e} \int dz dM \frac{(1+z)^2}{4\pi H(z)} \frac{dn}{dM} \int d\vec{r} P_e(|\vec{r}|; M, z), \quad (3.4)$$

where dn/dM is the halo mass function and \vec{r} denotes position within a given halo, and the expression assumes a flat universe. The expression for $\langle T_e \rangle_{\text{hm}}$ is analogous. Note that the halo model neglects the IGM contribution discussed in Appendix B.4.

In the following, it will be useful to think of the observables $x_i \equiv \{\langle y \rangle, \langle T_e \rangle\}$ in terms of the approximate factorization

$$x_i \sim f_i^c(\sigma_8, \Omega_m, \dots) f_i^b(\{A_j\}) f_i^{\text{CV}}(\delta), \quad (3.5)$$

where f_i^c describe the dependence on cosmological parameters, f_i^b are functions of a set of feedback parameters A_j , and f_i^{CV} depends on the initial conditions and thus encapsulates sample variance. Such factorizations are frequently-used and good approximations to observables that are well-described by the halo model. Nonetheless, our results typically only weakly depend on the validity of this approximation.

It should be noted that we group all our uncertainty on the simulation sub-grid model in the feedback parameters A_j . These supernova and AGN feedback parameters, further elaborated on in Sec. 3.2.1, predominantly affect the ICM contribution to the distortion signals. There is a non-negligible IGM contribution, however, which in Appendix B.4 we show is a $\sim 10\%$ effect with $\sim 40\%$ theoretical uncertainty.

The rest of this paper is structured as follows. In Sec. 3.2 we describe the CAMELS simulations as well as the larger reference boxes. Sec. 3.3 provides a short summary of the assumed experimental setup used for forecasting. In Sec. 3.4 we describe how we interpolate through the CAMELS data. Sec. 3.5 contains the main results of this work, namely, dependence of the distortion signals on feedback parameters and a Fisher forecast. We conclude in Sec. 3.6. The appendices contain several technical details and some new computations that did not fit in the main discussion.

3.2 Simulations

The y -distortion components described above can easily be measured from a hydrodynamical simulation. In fact, Eq. (3.1) can be rewritten as

$$\langle y \rangle = \frac{\sigma_T}{m_e} \int dz \frac{(1+z)^2}{H(z)} \langle P_e^c(z) \rangle, \quad (3.6)$$

where P_e^c is now in comoving units and the average is over the volume of a given simulation snapshot. An analogous expression holds for $\langle T_e \rangle$.

3.2.1 CAMELS

We primarily use the CAMELS suite of hydrodynamical simulations [77] that consists of several thousand $25 h^{-1}\text{Mpc}$ boxes, each run with 256^3 dark matter particles and 256^3 initial fluid elements. Each simulation is described by the following parameters: *i*) the simulation code/subgrid model, *ii*) two cosmological parameters (σ_8 , Ω_m), *iii*) four feedback parameters⁶ ($A_{\text{SN}1}$, $A_{\text{SN}2}$, $A_{\text{AGN}1}$, $A_{\text{AGN}2}$), and *iv*) the random seed for the initial conditions. The remaining cosmological parameters are fixed at *Planck*-compatible flat ΛCDM values, $\Omega_b = 0.049$, $h = 0.6711$, $n_s = 0.9624$ [8].

Two different simulation codes are used. The simulations labeled ‘IllustrisTNG’ were run with the Arepo code [122, 132] and the same subgrid model as the flagship IllustrisTNG simulations [133–138]. The simulations labeled ‘SIMBA’ were run with the GIZMO code [139] and the same subgrid model as the flagship SIMBA simulations [140]. These codes differ substantially in their subgrid implementations, so having comparable simulations with both gives a good indication of the theoretical uncertainty.

The cosmological parameters σ_8 , Ω_m are varied in the intervals $[0.6, 1.0]$ and $[0.1, 0.5]$, respectively, the fiducial model being $(0.8, 0.3)$.

The precise definition of the four feedback parameters is given in Ref. [77]. Broadly, $A_{\text{SN}i}$ parameterize the subgrid prescription for galactic winds, while $A_{\text{AGN}i}$ describe the efficiency of black hole feedback. The $i = 1$ components can be thought of ‘energy’ normalizations, while the $i = 2$ components scale the speed of outflows. In detail, however, the meaning of the feedback parameters differs substantially between IllustrisTNG and SIMBA. We thus caution against any direct comparison in terms of the feedback parameters between the two subgrid models. All the A_j are multiplicative factors relative to the fiducial efficiencies in the original IllustrisTNG and SIMBA simulations; the fiducial model is therefore unity for all feedback parameters. $A_{\text{SN}1}$ and

⁶Note that other works, e.g. Ref. [131], explicitly differentiate in their notation between the IllustrisTNG and SIMBA feedback parameters. We choose not to do so since it simplifies the notation in various places.

$A_{\text{AGN}1}$ are varied in $[0.25, 4.0]$, while $A_{\text{SN}2}$ and $A_{\text{AGN}2}$ are varied in $[0.5, 2.0]$. These intervals were chosen heuristically by the CAMELS team, and as we will see the corresponding variations in the y observables differ drastically between IllustrisTNG and SIMBA. Some of the simulations at the corners of parameter space are so extreme that they are certainly not realistic, for example with regard to galaxy properties.

For each of the two simulation codes, the CAMELS suite comprises the following sets of simulations, all of which will be used in this work.

- LH: (latin hypercube) 1000 simulations in which cosmology and feedback parameters are varied on a latin hypercube, each run having a different random seed,
- 1P: (one parameter at a time) 10 variations for each cosmological and feedback parameter individually at fixed random seed,
- CV: (cosmic variance) 27 simulations at the fiducial model with differing random seeds.

3.2.2 Larger boxes

In addition to the CAMELS suite, we also use larger boxes, namely IllustrisTNG300-1 ($205 h^{-1}\text{Mpc}$) and SIMBA100 ($100 h^{-1}\text{Mpc}$). These simulations are useful to calibrate against the fact that $\langle f_i^{\text{CV}}(\delta) \rangle$ (c.f. Eq. 3.5) is relatively more biased for the small $25 h^{-1}\text{Mpc}$ CAMELS boxes. Denoting an observable measured in one of the large boxes as x_i^{lb} (c.f. Eq. (3.5) for the notation), we compute an estimator for this multiplicative bias as

$$b_i = x_i^{\text{lb}} / \langle x_i \rangle_{\text{CV}}, \quad (3.7)$$

where the latter average is over the CAMELS CV set and we apply the correction formula

$$x_i \leftarrow x_i b_i \quad (3.8)$$

	$\log_{10} 10^6 \langle y \rangle$		$\log_{10} \langle T_e \rangle [\text{keV}]$		$10^6 \langle y \rangle$		$\langle T_e \rangle [\text{keV}]$	
	IllustrisTNG	SIMBA	IllustrisTNG	SIMBA	IllustrisTNG	SIMBA	IllustrisTNG	SIMBA
CAMELS CV	0.046 ± 0.026	0.369 ± 0.014	-0.386 ± 0.034	0.1407 ± 0.0054	1.11	2.34	0.41	1.38
large box	0.1928	0.3995	0.02728	0.2349	1.56	2.51	1.06	1.72
$\Delta[\sigma]$	-5.6	-2.1	-12	17				
b_i					1.41	1.07	2.59	1.25

Table 3.1: Comparison between statistics over the CAMELS CV set and the larger boxes of sizes $L = 205$ and $100 h^{-1}\text{Mpc}$ for IllustrisTNG and SIMBA respectively. The left half of the table is in log since log-normal is a decent approximation; for clarity the right half gives the same information in more familiar units but without error bars. The error bars in the first line are on the mean of the CV set. The error bars on the large box measurements are, in the naive Poissonian approximation $\sqrt{(L/25 h^{-1}\text{Mpc})^3/27}$, smaller by factors 4.5 and 1.5 for IllustrisTNG and SIMBA respectively, and omitted in the table. All large-box measurements have been rescaled to the fiducial CAMELS cosmology, as explained in Sec. 3.2.2. The third line lists the differences between the large-box and CV measurements, in units of the error bars on the CV means; these numbers highlight significant biases in the CV set due to box-size effects. The last line gives the corresponding bias factors introduced in Eq. (3.7).

to the $\langle y \rangle$, $\langle T_e \rangle$ measured in the CAMELS simulations. Values for the b_i are listed in Tab. 3.1.

The large boxes have the same subgrid prescription as the CAMELS fiducial model with the corresponding label, but slightly different cosmologies. We account for this fact by rescaling x_i^{lb} to the CAMELS fiducial cosmology, using power laws in h , Ω_m , Ω_b , n_s , σ_8 . These power laws were fitted using the halo model, since the scalings that could be derived from the 1P set are unreliable and do not encompass all differences in cosmology. We refer to Sec. 3.4.2 for a detailed description of the assumptions made in the halo model calculation. For reference, the power laws are listed in Appendix B.2.

There is, of course, a remaining bias due to the finite volume of both the larger boxes and the CV set. However, as seen in Tab. 3.1, this is a relatively small source of error compared to the overall bias, since Eq. (3.7) is significantly different from unity for all cases except perhaps the SIMBA $\langle y \rangle$. We note that the finite error bars on the large box measurements do not affect our forecasts in the later parts of this work.

3.3 PIXIE experimental model

For deriving our forecasts later in section 3.5, we use parameters corresponding to the ‘extended’ PIXIE experiment [98, 99, 141] from Ref. [78]. Their most complete model includes marginalization over a variety of foregrounds, namely galactic dust thermal emission, the cosmic infrared background, synchrotron radiation, free-free emission, spinning dust (anomalous microwave emission), and integrated CO. For each foreground component, fitting formulas were assumed for the spectral energy distribution (SED) according to *Planck* measurements [142]. There are some caveats to this approach, related to the spatial variation of the foregrounds and the relatively simplistic modeling of their SEDs [see 143, for related discussion]. However, for the purposes of this work, the derived forecasts should be accurate enough. The forecast considers all non-negligible CMB spectral distortion signals (blackbody, y , relativistic y, μ) and marginalizes over them when computing the $\langle y \rangle$ - $\langle T_e \rangle$ posterior.

While Ref. [78] used a full MCMC pipeline to arrive at their posteriors, in the $\langle y \rangle$ - $\langle T_e \rangle$ plane, these are well approximated as Gaussian. We thus compress the marginalized posterior into a simple 2×2 covariance matrix, illustrated in the inset in Fig. 3.1. In general, the fiducial models used in this work differ somewhat from the fiducial model of Ref. [78]; we assume that the covariance matrix is constant regardless of the mean. This approximation is not a dominant source of systematic uncertainty.

3.4 Interpolating neural networks

Using Eq. (3.6), we compute $\langle y \rangle$ and $\langle T_e \rangle$ for the CAMELS LH, CV, and 1P sets. It is our primary goal to extract the functions $f_i^b(A_{\text{SN}1}, A_{\text{AGN}1}, A_{\text{SN}2}, A_{\text{AGN}2})$, in the language of Eq. (3.5). It may be argued that in the limit in which these f_i^b factorize even further into functions of the individual feedback parameters (which, as we shall see, is quite a good approximation), the 1P set should be all we need. However,

the substantial sample variance in the small CAMELS boxes makes this approach unreliable. Therefore, we will instead utilize the LH set. There, σ_8 , Ω_m , and the four feedback parameters are varied by sampling from a latin hypercube. Furthermore, each data point is a noisy sample in a 6-dimensional space. Thus, we will use neural networks as smoothing interpolators through the LH set.

3.4.1 Training

We aim to learn functions

$$\langle y \rangle \text{ or } \langle T_e \rangle = F_\alpha(\sigma_8, \Omega_m, \{A_j\}), \quad (3.9)$$

where $\alpha \in \{\text{IllustrisTNG, SIMBA}\}$ and F is parameterized as a multi-layer perceptron. A multi-layer perceptron is a series of affine transformations $W\mathbf{x} + \mathbf{b}$, each followed by a non-linear activation function (in our case the leaky rectifying linear unit). To mitigate overfitting, we allow dropout, i.e. probabilistic zeroing of neurons during training.

The relatively small dataset of 1,000 LH simulations makes this a somewhat non-trivial task, thus we perform automated hyperparameter optimization using the Optuna package [144] to converge at good architectures. Generally, our networks have 2 – 4 hidden layers with a few hundred neurons each and high dropout rates 20 – 70 %. A useful null-test is the following. From the 1P set we can interpolate the dependence on σ_8 , Ω_m (simple linear interpolators which will generally be biased due to sample variance). Using these interpolants, we can remove most of the dependence on the cosmological parameters from the LH data (c.f. the cyan data points in Fig. 3.3). If the neural net is well-converged, it should matter very little whether the training data have the full dependence on cosmology or whether it is mostly factored out. The point here is that we are essentially fitting the same data, just with different transformations

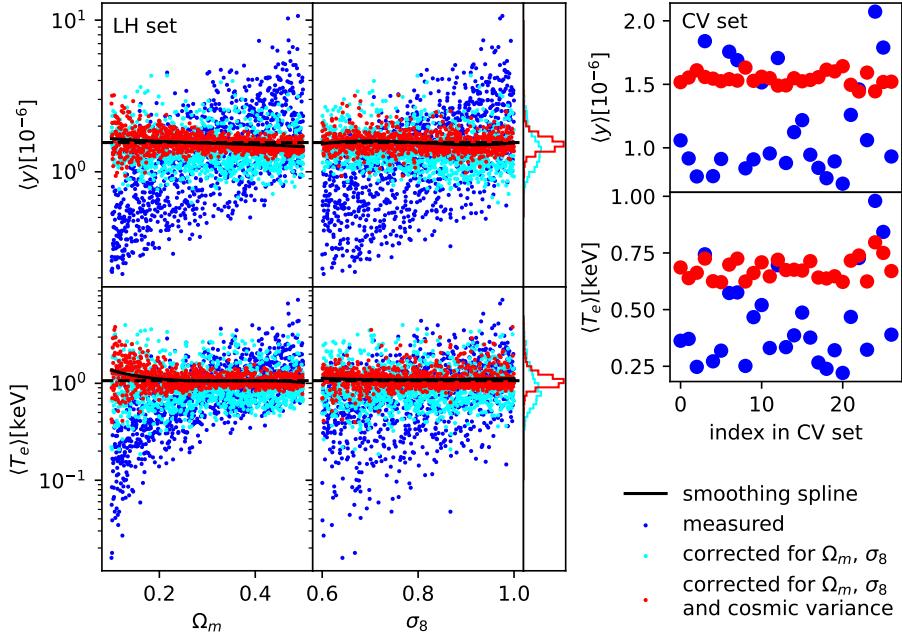


Figure 3.3: Illustration of the effectiveness of analytical correction factors for the IllustrisTNG simulation outputs. Blue markers are direct measurements from the simulations. The cyan markers represent a rescaling to the fiducial cosmology ($\Omega_m = 0.3$, $\sigma_8 = 0.8$), using fits to the CAMELS 1P set. Since the CAMELS CV set is at constant cosmological and astrophysical parameters, the cosmology correction is meaningless there. For the red markers, in addition to the cosmology correction, a correction for sample variance has been applied, as described in the text. The dashed black lines represent the ‘ground truth’ measurements from the larger boxes, while the solid black lines are smoothing splines through the red markers (these are solely to guide the eye). For the LH panels on the left, all data points have been rescaled using Eq. (3.7). We have also plotted histograms corresponding to the cyan and red markers.

applied. It is not actually very important what these transformations are as a function of σ_8 and Ω_m , we simply chose ones that remove most of the dependence on these cosmological parameters.

Indeed, in the case of SIMBA, this null test is passed satisfactorily. However, IllustrisTNG is more challenging, since the feedback parameters generally have a much smaller effect on the y observables in this simulation. Thus, we need to apply an additional transformation, as described in the following section.

3.4.2 Halo model correction factors for IllustrisTNG

As mentioned before, naively training the neural networks on the IllustrisTNG LH simulations does not yield robust results according to the null test described above. The reason is that, in contrast to SIMBA, the dependence on feedback parameters is obscured by the overwhelming noise due to sample variance. This issue can be seen in Fig. 3.1, where for IllustrisTNG the 1P data points scatter much less than the LH points. There is, however, a simple method to mitigate this problem. As we have argued in the introduction, the halo model provides a relatively good description of sparse fields such as Compton- y . Furthermore, the factorization of Eq. (3.5) should be a good initial guess. This motivates the following ‘correction’ formula:

$$\langle y \rangle \leftarrow \langle y \rangle \frac{\langle y \rangle_{\text{hm}}^{\text{Tinker HMF}}}{\langle y \rangle_{\text{hm}}^{\text{measured HMF}}} . \quad (3.10)$$

Here, ‘Tinker HMF’ stands for the halo mass function from Ref. [145] and the expression is analogous for $\langle T_e \rangle$. On the other hand, the mass function can also be measured in the individual CAMELS simulations. In order for the numerical integrations to be well behaved, we express it as

$$\frac{dn}{d \log M} \sim \exp \left[-\frac{1}{2} \left(\frac{\log M'}{\sigma} \right)^2 \right] * \sum_i \delta(\log M' - \log M_i), \quad (3.11)$$

where $*$ indicates convolution and we have dropped some prefactors. Our results are relatively insensitive to the hyperparameter σ ; in the following we will use $\sigma = 0.19$ dex. The halo masses M_i were measured using a friends-of-friends finder [146]. We interpolate the smoothed mass function in mass and redshift using a bilinear routine.

In order to perform the halo model calculation, we use the pressure profile fitting formula from Ref. [147] and assume isothermal halos with temperatures according to

Ref. [148]. We apply a 20 % correction to the halo masses entering the temperature fitting formula to account for hydrostatic mass bias; the same value was used in Ref. [85]. It should be noted that the fitting formula from Ref. [148] is likely not as accurate as more recent proposals [e.g. 87], but as we show in the following, it is sufficient for our purposes. Following Ref. [85], we assume a radial cut-off at $2.5 R_{\text{vir}}$ with the virial radius definition of Ref. [149].

Of course, this procedure rests on a number of assumptions, and we should verify that it yields reasonable results. Fig. 3.3 shows in the right panel in blue the original measurements and in red the mass function-corrected values for the 27 realizations in the CV set. Indeed, we see that the corrected values scatter much more tightly; the fact that the procedure is more successful for $\langle y \rangle$ than $\langle T_e \rangle$ is likely because the temperature fitting formula used is calibrated at larger halo masses than those that dominate the signal. In the left panel of Fig. 3.3, we show projections of the LH set onto the σ_8 and Ω_m axes. As expected, the blue measurement data points have a large scatter and show a mean evolution with the cosmological parameters. The cyan data points illustrate our procedure for dividing out most of the cosmology dependence, as described in the previous section. However, they still display a large scatter which is dominated by sample variance. Finally, the red data points have the mass function correction applied in addition. It is striking how much tighter they cluster compared to the cyan markers, again indicating that the mass function correction is working rather well. The scatter is relatively independent of cosmology, except for the low- Ω_m points where the halo model approach seems to start to break down. We observe that the remaining scatter in the LH panels significantly exceeds what is observed for the red markers in the CV panels, demonstrating that applying the mass function correction yields an LH set that is dominated by signal from the feedback parameters. Indeed, the neural networks trained on these de-noised LH simulations pass the null test described in Sec. 3.4.1.

It is worth noting that we tried the described mass function correction procedure for SIMBA as well. However, for these simulations it did not work well at all, indicating that halos in SIMBA are not well approximated by the fitting formulae we used in the halo model (this may be related to the very long-range effects feedback has in SIMBA [150]). Fortunately, the signal from feedback parameters is strong enough in SIMBA that it was possible to extract from the LH data with full sample variance contamination.

3.5 Results and Discussion

In the previous section, we have constructed interpolating neural networks which return the y distortion monopoles $\langle y \rangle$ and $\langle T_e \rangle$ as functions of astrophysical feedback parameters $A_{\text{SN}1}$, $A_{\text{AGN}1}$, $A_{\text{SN}2}$, $A_{\text{AGN}2}$ and cosmological parameters Ω_m , σ_8 . We now aim to explore the dependence on the astrophysical parameters and translate it into forecast constraints assuming the spectral distortion measurement described in Sec. 3.3.

3.5.1 Parameter dependence

We first explore how $\langle y \rangle$ and $\langle T_e \rangle$ depend on individual feedback parameters with the remaining ones (as well as σ_8 and Ω_m) fixed. In Fig. 3.4, we show as blue markers the measurements from the CAMELS 1P set, rescaled such that, their value at the fiducial point, is equal to that from the corresponding large box (indicated by the red crosses). The blue lines in Fig. 3.4 are evaluations of the trained neural nets. Agreement between the 1P results and the neural nets is generally better for SIMBA, because the observables depend much more strongly on the feedback parameters than in IllustrisTNG. For IllustrisTNG, there are substantial differences, although in many cases the qualitative trends are similar in 1P and the neural nets. As we have argued

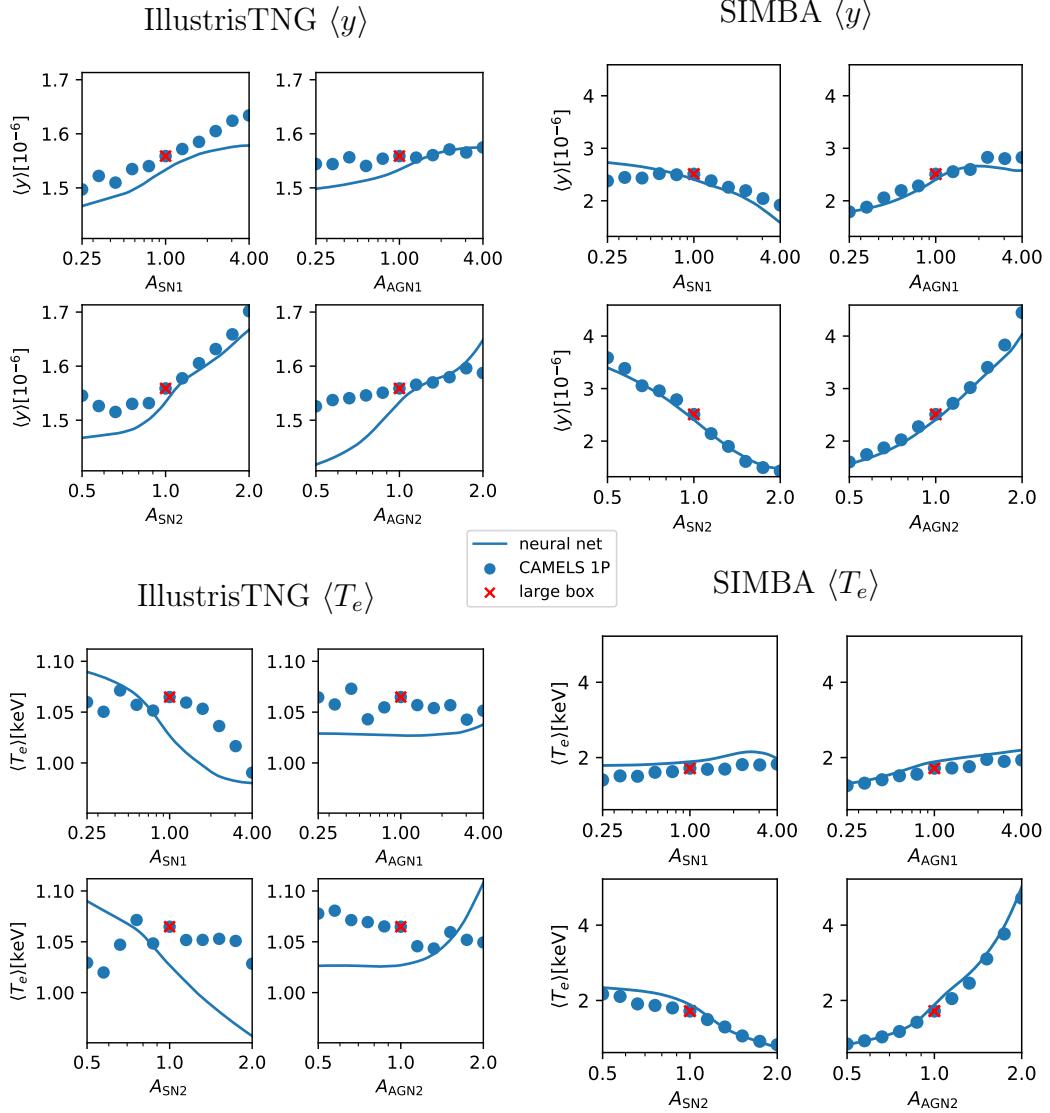


Figure 3.4: Dependence of spectral distortions on individual simulation feedback parameters. In each panel, all cosmological and astrophysical parameters are held fixed apart from the one shown on the horizontal axis. The blue markers are measurements from the CAMELS 1P set, rescaled such that at the fiducial model they agree with the measurements from the larger boxes, which are indicated by the red crosses. The blue lines are evaluations of neural networks trained on the CAMELS LH set. We argue in the text that the LH-trained networks are likely more robust than the 1P data in most cases. Note that the vertical plot ranges are identical within each quadrant.

before, we believe that generally the neural nets are more robust since they marginalize over the simulation initial conditions. However, this may not be true for some of the cases in IllustrisTNG for which the dependence is comparatively weak, because the neural net may have little incentive to accurately capture this dependence. There are, however, strong counterarguments against this hypothesis. It seems clear from the SIMBA panels that in almost all cases the SN and AGN parameter pairs each work in the same direction. This behaviour is mirrored in IllustrisTNG, also in the cases where strong deviations from the 1P set are observed. We emphasize that our neural net architecture has no implicit preference for grouping the parameters in such a way. Furthermore, we have evidence that the initial conditions for the 1P set happen to be somewhat anomalous. Visually, this manifests itself by the presence of an unusually large void in the $z = 0$ density field. More quantitatively, this can be seen in Fig. 3.1, where the 1P data points (for both IllustrisTNG and SIMBA) tend to fall at higher $\langle T_e \rangle$ than the mean trend in the LH set. Finally, as mentioned in Sec. 3.4.1, we have performed a non-trivial null-test on the trained networks which is passed satisfactorily (after correcting for sample variance in the case of IllustrisTNG). Thus, we believe that in fact for all cases the neural nets are closer to reality than the 1P set.

Turning now to physical interpretation, we observe striking differences between SIMBA and IllustrisTNG. In the case of $\langle y \rangle$, IllustrisTNG has positive slopes for all four feedback parameters, while for SIMBA the SN parameters exhibit negative slopes. It seems rather natural that slopes should generally be positive because integrated electron pressure is a measure of thermal energy. However, it has also been established that stronger supernova feedback can counteract the AGN contribution, since the outflow of matter due to stellar feedback limits the black hole growth rate [151] (see also [152]). Ref. [152] used the OWLS subgrid model which in terms of $\langle y \rangle$ and $\langle T_e \rangle$ is closer to SIMBA than to IllustrisTNG. This is consistent with the frequency of negative slopes with A_{SNi} in Fig. 3.4.

Interestingly, in some cases there are saddle points. That these are probably mostly real and not artifacts in the neural networks is supported by the fact that they are also visible in the 1P markers. Fortunately, these saddle points are generally away from the fiducial model (except for the AGN dependence of $\langle T_e \rangle$ in IllustrisTNG), so they will not contaminate the Fisher forecasts significantly.

As we show in Appendix B.5, the leading-order behaviour of the y -distortion observables as a function of the four feedback parameters is well-described by a factorization,

$$x_i = \prod_j g_i^{(j)}(A_j), \quad (3.12)$$

where the $g_i^{(j)}$ are identical to the 1-parameter functions in Fig. 3.4 up to normalization.

As a cross-check, we also test our results using another machine learning tool called symbolic regression. The results, presented in Appendix B.6 support the factorization approximation Eq. (3.12).

3.5.2 Fisher forecasts

We now turn to forecasting constraints on the simulation feedback parameters given our fiducial model for the PIXIE constraints on $\langle y \rangle$, $\langle T_e \rangle$ presented in Sec. 3.3. It should be emphasized that the primary goal of this exercise is to demonstrate the substantial constraining power of a realistic y -distortion measurement on our understanding of astrophysical feedback processes. In particular, the uncertainty on the fiducial model implies that the constraints presented in the following should only be interpreted as qualitative indicators. We make the approximation that the cosmological model (specifically Ω_m and σ_8) is known perfectly; in comparison to our ignorance on the subgrid model this is a good assumption.

As we have already seen before, the feedback parameters have very different consequences in IllustrisTNG and SIMBA. We will need to assume that the chosen

priors on these parameters are reasonable in terms of other observables, so that a direct comparison between the constraints (in units of the prior) makes any sense at all.

First, we construct the 4×4 Fisher matrices for IllustrisTNG and SIMBA as usual,

$$F_{ab} = \frac{\partial x^i}{\partial A^a} C_{ij}^{-1} \frac{\partial x^j}{\partial A^b}, \quad (3.13)$$

with C the 2×2 covariance matrix, marginalized over all foreground nuisance parameters (c.f. Sec. 3.3), and the derivatives are computed from the neural network interpolators with finite difference step sizes chosen to match the scale of the posterior. Of course, these Fisher matrices are degenerate since two measurements are not sufficient to constrain four parameters.

For each simulation type, by diagonalizing the Fisher matrix we identify the best-constrained orthogonal parameter combinations as

$$\begin{aligned} A_{\text{IllustrisTNG}}^{(1)} &= A_{\text{SN1}}^{+0.20} A_{\text{AGN1}}^{+0.13} A_{\text{SN2}}^{+0.77} A_{\text{AGN2}}^{+0.59}, \\ A_{\text{IllustrisTNG}}^{(2)} &= A_{\text{SN1}}^{+0.52} A_{\text{AGN1}}^{-0.10} A_{\text{SN2}}^{+0.44} A_{\text{AGN2}}^{-0.73}, \\ A_{\text{SIMBA}}^{(1)} &= A_{\text{SN1}}^{-0.17} A_{\text{AGN1}}^{+0.30} A_{\text{SN2}}^{-0.72} A_{\text{AGN2}}^{+0.61}, \\ A_{\text{SIMBA}}^{(2)} &= A_{\text{SN1}}^{+0.42} A_{\text{AGN1}}^{-0.37} A_{\text{SN2}}^{+0.36} A_{\text{AGN2}}^{+0.74}. \end{aligned} \quad (3.14)$$

Note that our methodology of finding the most constrained parameter combination(s) is similar to that used in galaxy photometric survey studies for identifying the $S_8 \equiv \sigma_8 \Omega_m^\alpha$ parameter combination [153, 154]. The induced 2×2 Fisher matrix on the thus identified subspaces tangent to the fiducial model is not singular. Of course, these combinations are not unique since the Fisher matrix only gives linear-order information. Not surprisingly given the intuition from the upper half of Fig. 3.4, in $A^{(1)}$ the powers have identical signs for IllustrisTNG, while for SIMBA the SN and AGN parameters have powers of opposite signs.

We compute the 68 % constraints

$$\begin{aligned}\sigma(A_{\text{IllustrisTNG}}^{(1)}) &= 0.015, & \sigma(A_{\text{SIMBA}}^{(1)}) &= 0.0024, \\ \sigma(A_{\text{IllustrisTNG}}^{(2)}) &= 1.3, & \sigma(A_{\text{SIMBA}}^{(2)}) &= 0.075.\end{aligned}$$

Thus, a PIXIE-like experiment could place percent-level constraints on parameter combinations that are currently only known to little better than an order of magnitude. In the case of SIMBA the measurement of the relativistic component would also allow a $\sim 10\%$ constraint on a second parameter combination. This is not possible for IllustrisTNG, due to the extremely small variation of $\langle T_e \rangle$. Conversely, this implies that a measurement of $\langle T_e \rangle$ significantly different from the fiducial model has the potential to simply rule out the CAMELS-based implementation of the IllustrisTNG model.

Robustness check

We have argued before (Sec. 3.5.1) that we believe the neural network interpolators to be more robust than the 1P data, particularly for IllustrisTNG where they disagree substantially. As a robustness check, we have repeated the Fisher analysis using derivatives estimated from one-dimensional interpolators through the 1P set. In the case of IllustrisTNG, the error bar on the best-constrained parameter combination inflates by a factor ~ 3 , while the orthogonal combination remains unconstrained. The slight inflation is due to the derivatives from the 1P set being slightly shallower than those from the neural net, see Fig. 3.4. In the case of SIMBA, the error bars increase by about 20 %. The corresponding parameter combinations differ from the ones listed in Eq. (3.14), but the qualitative features (relative magnitudes and signs of the exponents) are very similar. These comparisons indicate that our results for

IllustrisTNG are robust within a factor of a few while those for SIMBA are very accurate.

3.6 Conclusions

This work is one of the first systematic studies of the information content in the late-time Sunyaev-Zel'dovich spectral distortions. Besides a simulation-based forecast in the context of specific subgrid models, in the appendices we have also given some novel theoretical results regarding the IGM and reionization contributions.

We have measured the non-relativistic and relativistic mean spectral distortion amplitudes $\langle y \rangle$ and $\langle T_e \rangle$ in the CAMELS simulations suite. By training neural networks on this data, we have constructed interpolators returning the two signals as a function of Ω_m , σ_8 , and four feedback parameters. In the case of IllustrisTNG, the observables depend only weakly on feedback, necessitating the use of halo model-derived correction factors to reduce the large sample variance due to the small CAMELS box size.

Incidentally, the described method to reduce scatter arising from small simulation boxes by using mass function-dependent scaling factors should be more generally applicable to many works concerned with fields that can be approximated with a halo model. We have also tried to reduce biases as much as possible by matching our data to the comparatively much larger size flagship IllustrisTNG and SIMBA simulations.

Using the interpolating neural networks, we have performed a Fisher forecast assuming a Gaussian posterior on $\langle y \rangle$ and $\langle T_e \rangle$, which had been computed assuming a PIXIE-like experiment and a realistic foreground model. Of course, our work is relevant for experiments other than PIXIE, e.g. ESA's Voyage 2050 large-scale proposal (see e.g. [101]), which is expected to reach even tighter constraints on distortion parameters.

We find that in the case of IllustrisTNG, only a single parameter combination can be strongly constrained, at the $\sim 2\%$ level. On the other hand, in the case of

SIMBA, the availability of two data points enables two orthogonal combinations to be measured, to $\sim 0.2\%$ and $\sim 8\%$. We emphasize again that although the feedback parameters are qualitatively similar between IllustrisTNG and SIMBA, their detailed meaning differs substantially so direct comparisons must be carried out with caution.

Given the limited information content from only two observables, it is interesting to ask what other measurements could be added in order to improve the constraints presented here. The tSZ and kSZ profiles of halos are already being used in order to assess the viability of simulation models, and Ref. [155] demonstrates using CAMELS that next-generation CMB experiments could place constraints on the feedback parameters. Similarly, Ref. [131] shows that deviations from self-similarity in the integrated tSZ flux – halo mass relation ($Y_{\text{SZ}} - M$) can also be used to constrain feedback parameters. A combination of such constraints could substantially improve upon the error bars presented in this work. Furthermore, the CAMELS feedback parameters also affect observables beyond the SZ effects, like the properties of galaxies. Typically observations of such astrophysical nature are difficult to propagate into hard posteriors, but at least qualitatively the check for simultaneous viability of a given subgrid model should be very useful for simulators.

The main source of uncertainty in the Fisher forecast is likely due to errors in the interpolators. We have performed extensive consistency checks on the trained neural nets, but the limited number of data points in the 1,000 CAMELS LH simulations places fundamental limits in the possible accuracy. For this reason, the given constraints have some associated uncertainty, of order unity for IllustrisTNG and at the 10% level for SIMBA. Nonetheless, our results are indicative of the transformative effect that measurements of the low-redshift spectral distortions could have for our understanding of baryonic feedback.

Acknowledgements. LT thanks Will Coulton, Shy Genel, and David Spergel for useful discussions. DW gratefully acknowledges support from the Friends of the Institute for Advanced Study Membership. The Flatiron Institute is supported by the Simons Foundation.

JCH acknowledges support from NSF grant AST-2108536. DAA was supported in part by NSF grants AST-2009687 and AST-2108944. JC was supported by the Royal Society as a University Research fellow (No. URF/R/191023) and by the ERC Consolidator Grant *CMBSPEC* (No. 725456). NB acknowledges support from NSF grant AST-1910021 and NASA grants 21-ADAP21-0114 and 21-ATP21-0129.

Chapter 4

Teaching neural networks to generate Fast Sunyaev Zel'dovich Maps

Abstract. The thermal Sunyaev-Zel'dovich (tSZ) and the kinematic Sunyaev-Zel'dovich (kSZ) effects trace the distribution of electron pressure and momentum in the hot Universe. These observables depend on rich multi-scale physics, thus, simulated maps should ideally be based on calculations that capture baryonic feedback effects such as cooling, star formation, and other complex processes. In this paper, we train deep convolutional neural networks with a U-Net architecture to map from the three-dimensional distribution of dark matter to electron density, momentum and pressure at ~ 100 kpc resolution. These networks are trained on a combination of the TNG300 volume and a set of cluster zoom-in simulations from the IllustrisTNG project. The neural nets are able to reproduce the power spectrum, one-point probability distribution function, bispectrum, and cross-correlation coefficients of the simulations more accurately than the state-of-the-art semi-analytical models. Our approach offers a route to capture the richness of a full cosmological hydrodynamical simulation of galaxy formation with the speed of an analytical calculation.¹

4.1 Introduction

Over the past decade, observations of the cosmic microwave background (CMB) have become one of the most powerful tools for studying the early Universe and for determining its basic properties (age, density, composition). Microwave background

¹based on Ref. [156], joint work with F. Villaescusa-Navarro, D. N. Spergel, D. Nelson, and A. Pillepich.

observations also have the potential to provide detailed quantitative measurements of the properties of the Universe’s recent evolution. Through the thermal Sunyaev-Zel’dovich (tSZ) effect, the microwave observations trace the large-scale distribution of electron pressure; while the kinematic Sunyaev-Zel’dovich (kSZ) effect traces the large-scale distribution of electron momentum [73, 74].

High resolution microwave background experiments such as ACT², SPT³, and upcoming instruments like the Simons Observatory⁴ and CMB S4⁵, provide ever improving measurements of small-scale fluctuations. Accurate theoretical predictions are essential for extracting the full information from these rich observations. Generating the simulations needed for these accurate predictions is particularly challenging for the SZ effects as they depend not only on the underlying cosmology but also on complex multi-scale processes including feedback from star formation and Active Galactic Nuclei and the rich plasma physics of cluster gas. Hydrodynamical simulations of the Sunyaev-Zel’dovich effects began before the establishment of precision cosmology [157–160]. Later works investigated the effects of sub-grid physics [161–165]. Large-scale simulation efforts have enabled direct comparison to microwave observations [166–172]. Hydrodynamical uniform-resolution simulations are computationally expensive and must adopt a trade-off between large volumes, with robust statistics of high-mass objects, and high numerical resolution to better capture the important physical processes.

Cosmologists have been taking a variety of approaches to more efficiently generate predicted tSZ maps. Most work builds on numerical simulations. Early papers assumed that on large scales the gas pressure traces the dark matter distribution [173, 174]. In order to more accurately describe the non-linear evolution, halo model approaches have been developed, starting from [175] and [176]. These assume that

²<https://act.princeton.edu>

³<https://pole.uchicago.edu>

⁴<https://simonobservatory.org>

⁵<https://cmb-s4.org>

the halos have a characteristic profile [177], depending on mass and redshift and possibly a few other parameters. A variety of other analytical approaches in the halo model framework have been developed [178–186]. These have some free parameters that are calibrated off observations of low-redshift objects. A comparison of different analytical halo model approaches is performed in [187]. Recent work has calibrated these halo electron pressure profiles off of numerical simulations [188, 147, 189–193], or observations [194–197, 105, 198–200]. A common theme of these semi-analytical models is the assumption of symmetries (often spherical) and the restriction to a small set of variables describing a given halo. Especially the neglect of halo sub-structure leads to errors in summary statistics [147].

Treatments of the kSZ effect begin with linear theory [201–203], and then were extended to non-linear scales [204–207]. More recent works use numerical simulations to calibrate the analytic treatment [208–210].

In this work we use a different approach, schematically illustrated in Fig. 4.1. We employ deep learning techniques to find the mapping between the spatial distribution of dark matter⁶ from gravity-only cosmological simulations and (1) electron pressure, (2) electron density, and (3) electron momentum from the corresponding state-of-the-art full-physics hydrodynamical realizations. In particular, we use simulations from the IllustrisTNG project [133–137, 172]. The advantage of this method with respect to the semi-analytical models is that gravity-only simulations are accurate enough to model the clustering of matter and halos down to rather small scales, and deep learning can transform their output to account for baryonic effects, capturing non-gravitational processes while avoiding oversimplifications such as spherical symmetry.

Machine learning has previously been used for similar tasks, for example to generate two-dimensional maps of the tSZ effect [211], and populating gravity-only simulations with galaxies [212–216].

⁶For brevity, here and in the remainder of this paper we use the phrase “dark matter” to mean the matter field obtained in a gravity-only simulation.

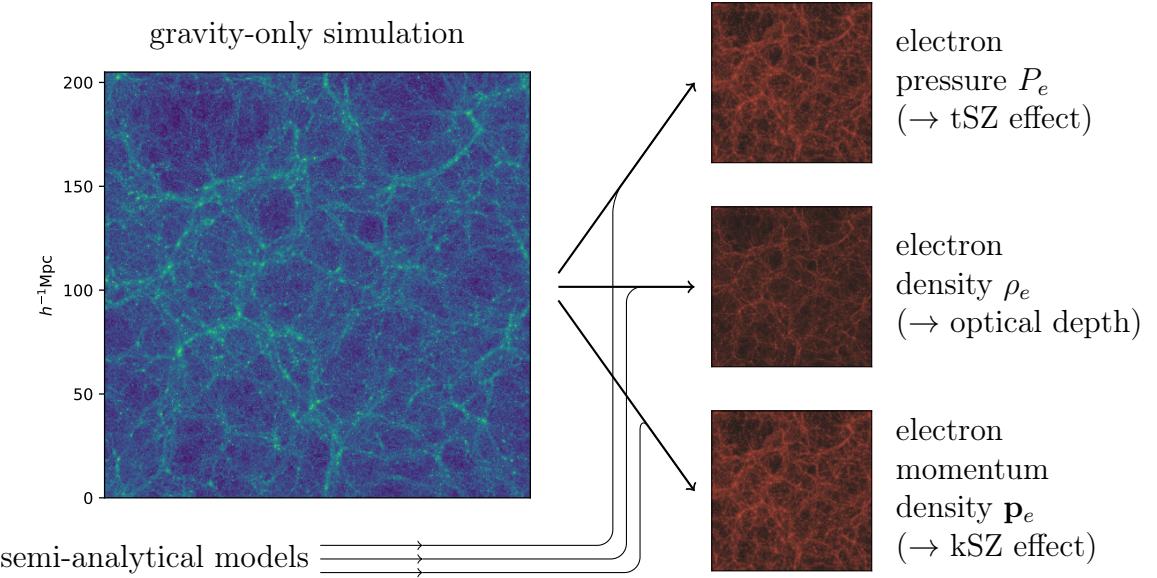


Figure 4.1: Schematic illustration of the approach taken in this work. We use properties of the matter field from gravity-only cosmological simulations, on the left, as input into convolutional neural networks which predict the electron gas properties via training based on the corresponding full-physics hydrodynamical realization, on the right. The networks utilize simulation-calibrated semi-analytical models for the targets, so that only residuals with respect to the hydrodynamical simulation output are required to be learned.

We approach this work as part of a broader program consisting of predicting baryonic observables from gravity-only simulations, or even directly from the initial conditions [217]. One application of particular interest would be the prediction of cross-correlation statistics between different observables. Because of our focus on cross-correlations, we are primarily interested in generating the low-redshift kSZ signal and defer the discussion of the high redshift signal to future work. Our current approach assumes that there is an injective map between the matter field and the observables of interest; we discuss stochasticity (which is not captured in our approach) in Section 4.6.

This approach will allow us to quickly generate tSZ and kSZ maps, over large areas of the sky, with the astrophysical model of the IllustrisTNG simulation [218, 219], but in comparatively negligible time (see the figures quoted in Section 4.6). Those

maps can be used to extract cosmological and astrophysical information from CMB observations alone, but also allow the modeling of cross-correlations such as tSZ-weak lensing or kSZ with spectroscopic surveys.

This paper is organized as follows. We describe the physics of the tSZ and kSZ effects in Section 4.2. Section 4.3 outlines the methods used in the analysis. Section 4.4 describes the challenge introduced by sparsity: rare massive clusters contribute a significant fraction of the tSZ signal and dominate the higher moment statistics of the maps. The section describes the use of zoom-in simulations of high density regions to augment the data set and ameliorate this issue. We show the results of our work in Section 4.5. Finally, in Section 4.6 we draw the conclusions. Some technical details are collected in the Appendices.

4.2 Physical background

At low frequencies, clusters cast shadows against the microwave sky: in the 90 and 150 GHz Planck, ACT and SPT maps, the tSZ effect makes clusters appear as cold spots. This effect arises from random motions of thermal electrons which scatter CMB photons into higher energy states. It produces a non-blackbody distortion in the primary CMB's Planckian spectrum, yielding a brightness temperature decrement (increment) at low (high) observation frequencies ν . It is usually parameterized in terms of the dimensionless Compton- y parameter, given by

$$y = \sigma_T \int_{\text{LOS}} dr \frac{P_e}{m_e}, \quad (4.1)$$

where $\sigma_T = (8\pi e^4/3)m_e^{-2}$ is the Thompson cross-section for electron-photon scattering, r is physical distance along the line of sight (LOS), $P_e = n_e T_e$ is electron pressure, and m_e is the electron mass. The dependence on the choice of line of sight \mathbf{n} has been

suppressed. Thus, the tSZ effect is a measure of the line-of-sight integrated thermal energy density in electrons.

The kinematic SZ (kSZ) effect, on the other hand, is sourced by coherent motion of the electrons. To leading order it preserves the blackbody energy distribution, yielding a shift in the local CMB temperature. We can write down a parameterization similar to the tSZ Compton- y , referred to as the Doppler- b parameter:

$$\mathbf{b} = \sigma_T \int_{\text{LOS}} dr \frac{\mathbf{p}_e}{m_e} e^{-\tau(r)}, \quad (4.2)$$

where

$$\tau(r) = \sigma_T \int_0^r dr' \frac{\rho_e}{m_e} \quad (4.3)$$

is the optical depth (average number of Compton scattering events up to r) and \mathbf{p}_e is the electron momentum density (momentum per unit volume).

For reference, the dimensionless parameters introduced above are related to the observed shifts in the CMB temperature by

$$\frac{\Delta T_{\text{tSZ}}(\mathbf{n}, \nu)}{T_{\text{CMB}}} = y(x \coth x/2 - 4), \quad x \equiv \nu/T_{\text{CMB}}; \quad (4.4)$$

$$\frac{\Delta T_{\text{kSZ}}(\mathbf{n})}{T_{\text{CMB}}} = -\mathbf{b} \cdot \mathbf{n}. \quad (4.5)$$

In the late-time universe, both SZ effects are predominantly sourced by electrons bound to dark matter halos. The tSZ effect is mostly sourced by the hot electron gas in the cluster center (due to the dependence on electron temperature), and the cluster integrated signal scales roughly as $Y \propto M_{\text{halo}}^{5/3}$ [74]. In contrast, the kSZ effect does not receive this temperature bias and is thus more sensitive to the cool electron gas in cluster outskirts, lower mass halos, and the IGM.

As we explain in Section 4.3.1, our procedure begins with semi-analytical models for the desired electron gas properties. For the electron pressure, we make use of the

simulation-calibrated model from [147], hereafter B12, which gives a simple fitting function for the electron pressure as a function of halo mass and redshift:

$$P_e^{\text{model}}(\mathbf{x}) = \sum_{\text{halos } h} \text{B12}(M_h, |\mathbf{x} - \mathbf{x}_h|). \quad (4.6)$$

The B12 semi-analytical model is calibrated on a particular cosmology with given assumptions on sub-grid baryonic effects.

In contrast to the electron pressure, we do not use a halo model fitting function [220] for the electron density, as we find that a simple linear fit relating dark matter and electron density with an additional Gaussian smoothing is good enough for the purposes of this work:

$$\rho_e^{\text{model}}(\mathbf{x}) = A \times [\text{Gaussian}(\sigma = 80 h^{-1}\text{kpc}) \circledast \rho_m](\mathbf{x}), \quad (4.7)$$

with A a scalar that was fitted to the TNG300 simulation. The convolution kernel was chosen by visually inspecting some halos, its precise shape and width are of little importance.

From this, we finally construct our semi-analytical model for the electron momentum density:

$$\mathbf{p}_e^{\text{model}}(\mathbf{x}) = \rho_e^{\text{model}}(\mathbf{x}) \mathbf{v}_m(\mathbf{x}), \quad (4.8)$$

where \mathbf{v}_m is the dark matter velocity field. In particular, we do not include any velocity bias between dark matter and baryon velocities.

4.3 Methods

In this section, we describe the methods used to prepare the data, as well as the general set-up for the machine learning procedure (input and target data, network

architecture). We defer the solutions we developed for specific challenges arising from the properties of the data to the next section.

4.3.1 Data preparation

For the majority of this work, we use the TNG300-1 (hydrodynamical) and TNG300-1_DM (gravity-only) $z = 0$ snapshots [133–137, 172]. We also make use of additional zoom-in hydrodynamical simulations, as discussed in Section 4.4.1; using an available subset of halos from the TNG-Cluster sample [221]. Details on the pixelization of the various fields are given in Appendix C.1.

In addition to the dark matter properties, we also use simple semi-analytical models for the output as input data. The reason for this is that neural networks train faster on residuals than on the entire target. We defer the reader to the previous section for a description of how these semi-analytical models have been constructed for the different targets.

4.3.2 Sampling

We expect the mapping between the dark matter field and the target properties of the electron gas to be dominated by operators whose spatial support does not exceed length-scales of a few Mpc. Therefore, we choose our training samples as local boxes, as described in more detail in Appendix C.2.

4.3.3 Neural network

Having defined the input and target data, we now discuss the neural network architecture. Given the locality of the problem as well as translational, rotational, and parity invariance, a deep convolutional neural network architecture is the logical choice. Motivated by previous successes in similar applications [e.g., 217, 213], we choose a

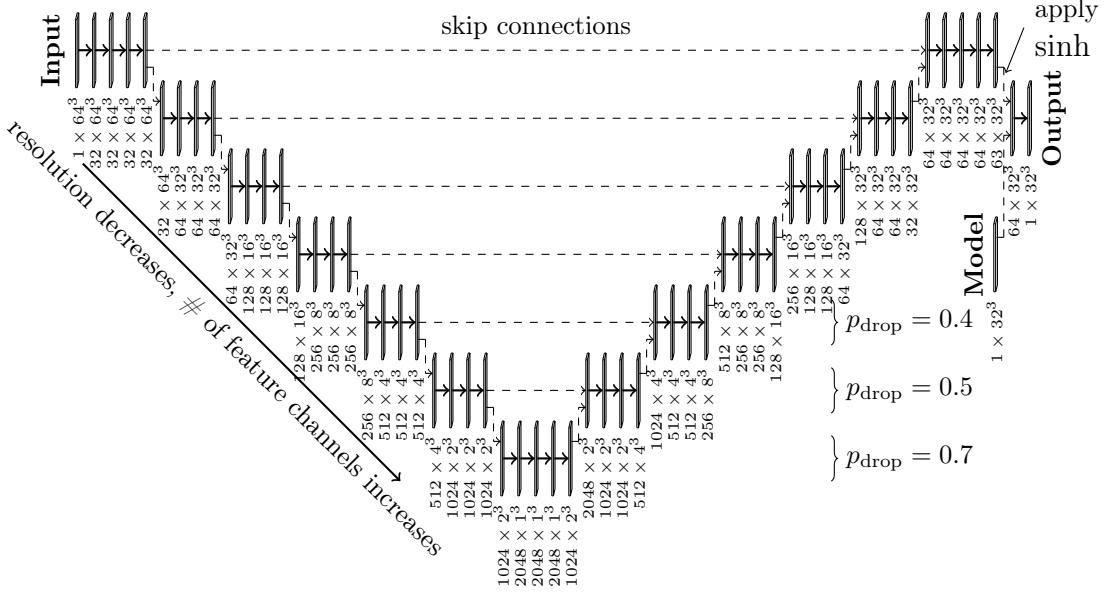


Figure 4.2: U-Net architecture for the CNN used in predicting electron pressure and density from dark matter distributions in 3D. The layer dimensions are given in the format (# of feature channels) \times grid size. The dashed lines represent copies, while the solid arrows are convolutional operations with 3^3 -sized filters (except for the very last operation, where the channels are simply collapsed into the output). If more than one dashed line ends at a single layer we concatenate along the feature channel direction. In the lowest three levels we apply dropout of entire feature channels with the probabilities given. We usually apply batch normalization, except wherever dropout is used. All activation functions are ReLUs. The network used to predict the electron momentum density is of similar structure, with the following differences: the semi-analytical model is not concatenated but rather multiplied with the first channel of the network representation; no sinh function is applied; since the resolution is a factor 2 smaller the network is smaller; in the two levels with highest resolution we apply Hardshrink activation functions. Most of these modifications are natural consequences of the negative-positive symmetry of the momentum data.

U-Net architecture, displayed in Fig. 4.2 (this network was used for electron pressure and density as target data, the architecture for the electron momentum density was slightly modified to take the sign symmetry of the data as well as lower resolution into account). The general idea of the U-Net is that while the receptive field increases in deeper layers, the number of feature channels increases. This allows the network to find more spatial correlations on larger scales. Skip connections (horizontal dashed lines in the figure) make it easier for the network to retain more local, high resolution features. Note that the semi-analytical model is fed into the network at a rather late stage, this makes it easier for the network to recognize that the target is close to the semi-analytical model with small residuals that need to be learned in the deeper layers. More details on the network architecture may be found in the caption of Fig. 4.2.

4.3.4 Training procedure

To train the network, we use the Adam optimizer with $\vec{\beta} = \{0.9, 0.999\}$ and choose samples from the training set according to a strongly biased selection function, as described in Section 4.4.1. These samples consist of in- and output boxes which are described in Appendix C.2 (there, we also specify the splitting into training, validation, and testing data). Although there is no natural definition of an epoch, for convenience, we define an epoch as 8192 training samples. The validation loss is evaluated on a random subset (chosen according to the same selection function as for the training set) of 256 boxes from the validation set. We find that a batch size of 32 yields the best performance. We generally start training with a learning rate of 10^{-3} , train for about 100 epochs, and then resume training from the state of the network with the best validation loss. We adjust the learning rate whenever we see the validation loss to become dominated by noise over any perceptible downwards trend. While during training only the loss function discussed in Section 4.4.1 is computed, after each training run (i.e. every ~ 100 epochs), we evaluate the network

on the whole validation box and compute power spectrum, cross-correlation coefficient with the reference hydrodynamical simulation, and one-point PDF. We base our decisions regarding the tuning of hyperparameters, changes in network architecture, and the eventual stop of training principally on these summary statistics as well as the validation loss curve.

We perform the usual data augmentation of rotations, reflections, and transpositions; in addition, in order to avoid overfitting, we multiply input data and target by voxel-wise Gaussian noise of 10 % standard deviation and unit mean. Although it is impossible to explore all directions in the space of hyperparameters, we are confident that with about 200 runs with different hyperparameters and network architectures we have found a relatively well optimized training procedure and network architecture.

4.4 Challenge: sparsity

The main challenge we need to overcome to train the neural network can be summarized as the sparsity of the dataset. In this section, we will argue that this problem comes in two different, but related, aspects, and explain the strategies we use to solve it. We will concentrate on the case of electron pressure here, the arguments carry over to electron density and electron momentum density.

4.4.1 Few interesting voxels

The first aspect to consider is the fact that only very few voxels are actually “interesting”, in the sense that they contribute to any commonly used summary statistics. This point is illustrated in Fig. 4.3, where we plot the electron pressure power spectrum. While the solid blue line is the power spectrum for the target TNG300 hydrodynamical box, the dashed lines result from setting voxels with electron pressure below varying thresholds to zero. We observe that only voxels with $P_e \gtrsim 10^{-2}\sigma(P_e)$ are relevant

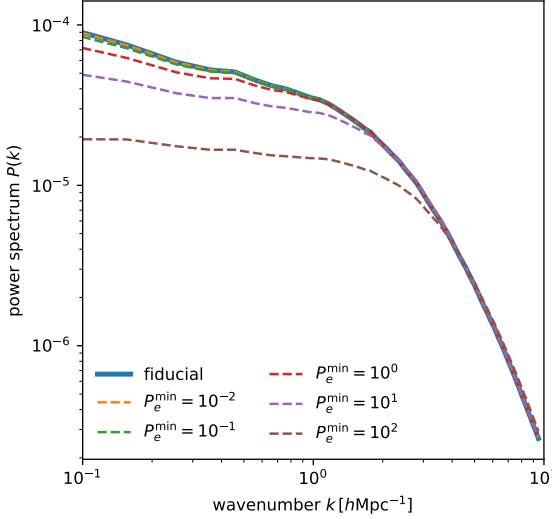


Figure 4.3: *solid blue*: electron pressure power spectrum for 20% of the TNG300 box. *dashed*: the same with electron pressure values below minimum values P_e^{\min} (given in units of $\sigma(P_e)$) set to zero. It can be seen that the electron pressure power spectrum can be reproduced using only voxels with $P_e \gtrsim 10^{-2}\sigma(P_e)$. These voxels represent only a small fraction of the entire box, as illustrated in Fig. 4.4.

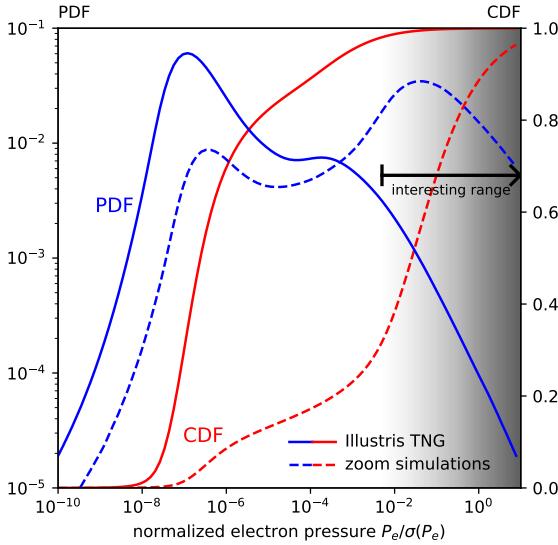


Figure 4.4: Illustration of the sparsity problem in the electron pressure distribution. Plotted are both probability distribution function (PDF; blue) and cumulative distribution function (CDF; red) for the TNG300 volume (solid) and the 181 TNG-Cluster zoom-in simulation boxes (dashed). The shaded area labeled “interesting range” was chosen according to the conclusions from Fig. 4.3. In the reference uniform-volume simulation, TNG300, only of order 1 % of the voxels fall into the “interesting” range.

(this argument is only rigorous for the case of the power spectrum, but it is clear that for higher order statistics the situation is even more severe). Referring now to the solid lines in Fig. 4.4, in which we plot the probability distribution function (PDF) and cumulative distribution function (CDF) of the electron pressure, we observe that those relevant voxels constitute only a tiny fraction (of order 1 %) of the simulation volume.

The small fraction of interesting voxels implies that a naive training procedure in which we would randomly pick training samples from the simulation volume would be

highly inefficient. Thus, we solve this first aspect of the sparsity problem by biasing the training samples. We know that the high electron pressure voxels are found in massive halos. Therefore, it seems natural to bias the sample selection such that always at least one halo of a given minimum mass is to be found in any training sample.

In the case of the electron density as target, we find that biasing according to the selection function⁷

$$w_{\rho_e}(M_{500c}) \propto \Theta(M_{500c} - 10^3)(\log_{10} M_{500c})^{0.25} \quad (4.9)$$

with Θ the Heaviside function gives good results.

Note that the actual halo mass function of the training samples is *not* identical to Eq. (4.9) but rather equals $w(M)$ times the TNG300 mass function (since $w(M)$ chooses from the available halos).⁸

In the case of the electron momentum density, we find that a less aggressive selection function works best:

$$w_{\mathbf{p}_e}(M_{500c}) \propto \Theta(M_{500c} - 10^3). \quad (4.10)$$

In the case of electron pressure, the biasing schemes described above are not sufficient. The reason is that the integrated electron pressure in halos of mass M_{halo} is not simply $\propto M_{\text{halo}}$, but scales as $\propto M_{\text{halo}}^{5/3}$. To overcome this problem, for electron pressure alone, we do not use the original TNG300 simulation for training samples, but instead draw our training samples from additional zoom-in simulations centered

⁷From here onwards, wherever not otherwise stated, we use the following units: unit mass = $10^{10} h^{-1} M_\odot$; unit length = $1 h^{-1} \text{kpc}$; unit speed = 1km/s .

⁸To be precise: we make a list of TNG300 halos with masses above the argument in the Θ -function, assign weights to them according to $w(M)$ and then, in each training step, randomly choose a halo according to these weights. The input (gravity-only) sample box, whose geometry is described in Appendix C.2, is then chosen such that the halo's center falls into one of its voxels, with equal probability for each voxel to contain the halo center.

on massive halos. These are performed with the same astrophysical galaxy-formation model as TNG300 and at similar resolution, but focus on high-mass halos [221]. Hence, these zoom-in simulations have a very different halo mass function compared to the TNG300 volume; in particular, while the TNG300 simulation only has a few halos with $M_{200c} > 10^{14.5} h^{-1} M_\odot$, the majority of the zoom-in primary objects are more massive than this. The electron pressure PDF and CDF obtained in these highly biased training samples are plotted as the dashed lines in Fig. 4.4. It can be seen that about 70 % of the voxels from the TNG-Cluster training set fall in the “interesting” range, according to the conclusions derived from Fig. 4.3. We point out that for the zoom-in simulations we did not have gravity-only simulations available. This implies that the mapping from input (which is gravity-only during validation and testing) to output is not exactly identical for the TNG-Cluster training samples and the TNG300 validation and testing set. We would expect this effect to be most prominent on small scales.

4.4.2 Tailed distributions

The second aspect of the sparsity problem is the remaining high dynamic range of the voxel values. Even though we have argued that the network can essentially ignore any $P_e < 10^{-2}\sigma(P_e)$, the remaining “interesting” electron pressure values still span several orders of magnitude (c.f. Fig. 4.4). Likewise, the dark matter density which we take as the network input also has a tailed distribution.

The solution for the problem of the tailed distribution of input data is relatively standard [e.g., 211], and consists in shrinking the dynamic range of the input data (this can be interpreted as a generalization of the unit-variance zero-centred rescaling that is standard practice in many machine learning applications). We use the transformation

$$\tilde{x}_{\text{DM}} = a[\log(1 + bx_{\text{DM}}) - c], \quad (4.11)$$

where b is related to the standard deviation of the field, and a and c are obtained by requiring zero mean and unit variance. For reproducibility, we list the values of these parameters in Appendix C.3.

The tailed target distribution constitutes a challenge in two ways. First, vanilla neural networks are not designed for the flow of high-dynamic-range data; standard architectures and training procedures are calibrated for well behaved distributions of the internal representation. To address this problem, we take two steps: (1) as mentioned before, we provide a semi-analytical model for the target, so that the network only has to learn the difference between target and semi-analytical model⁹; (2) before concatenating the network output with the semi-analytical model, we apply the sinh function to the network output, as illustrated in Fig. 4.2. We empirically found a sinh to work better than the exponential function, presumably the network prefers the possibility that some of the output data fall into the approximately linear range of the sinh. Both measures serve to decrease the dynamic range of the internal representation¹⁰.

The second challenge stemming from the tailed target distribution is related to the choice of loss function. It is clear that for good performance on summary statistics such as the power spectrum we would like to use a loss function of the form $L(p, t) = \Sigma(p - t)^2$, where p and t are the prediction and target respectively, and Σ schematically indicates summation over voxels. However, empirically it is found that starting training with this loss function leads to unstable behaviour, presumably because large gradients are backpropagated due to the high dynamic range of t .

⁹As can be seen from the network architecture, Fig. 4.2, the network has the freedom to rescale the semi-analytical model by a constant factor. In principle it could set this constant to zero if it deems the semi-analytical model too inaccurate, but empirically we find it to be close to unity. Thus, the word “residual” is to be understood in a somewhat generalized sense.

¹⁰In fact, we examined the values of weights and biases after training and found them to be approximately Gaussian distributed around mean zero with variance of order unity, which confirms the effect we expect the mentioned techniques to have.

Thus, in order to stabilize training, we adopt a time-dependent loss function,

$$L_\tau(p, t) = \sum [f_\tau(p) - f_\tau(t)]^2 ; \quad (4.12)$$

$$f_\tau(x) = e^{-\tau/\tau_0} \text{sgn}(x) \log(1 + 10|x|) + (1 - e^{-\tau/\tau_0})x , \quad (4.13)$$

where p and t have been divided by the standard deviation of the target field. The parameter τ measures time in units of epochs, and we found that a good choice for τ_0 is about 30-100, with lower values leading to more unstable training, while larger values prolong the training process unnecessarily and increase the risk of overfitting. The transformation f_τ smoothly interpolates between the more gentle $\log(1 + x)$ and a linear function, converging at the desired mean squared error (MSE) loss as training progresses beyond τ_0 epochs. The signum function and absolute value are only required for the electron momentum density which can have positive and negative values. We always evaluate the validation loss with $\lim_{\tau \rightarrow \infty} L_\tau$, i.e. the traditional MSE, since it is a good measure to quantify performance on the power spectrum and other commonly used summary statistics.

4.5 Results

Having explained the methods, we now proceed to present our results; we will split this section into four parts: In Section 4.5.1, we present the results for the electron pressure (i.e. tSZ effect) as target, Section 4.5.2 and 4.5.3 are concerned with the electron density and momentum density respectively, and in Section 4.5.4 we consider cross-correlations between different fields. We will give somewhat more detail on the electron pressure, with the results generalizing to a large extent to the other two cases. As usual in machine learning, the results here are evaluated on a testing set, i.e. parts of the simulation the network has not seen before, as described at the end of Appendix C.2. We will compare the performance of the network with that

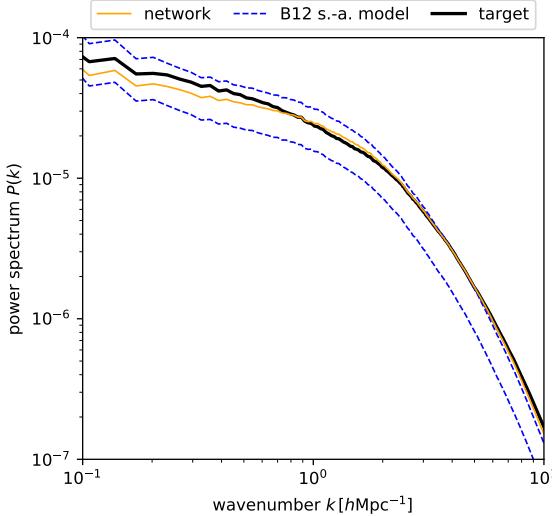


Figure 4.5: Electron pressure power spectrum. We plot two versions of the B12 semi-analytical model: the lower curve is the original output, the upper curve is the same rescaled by a constant factor to make the comparison fairer. While we observe excellent agreement between network prediction and target result (from the TNG300 hydrodynamical simulation), there is a lack of power ($\sim 20\%$) on large scales. This is likely related to errors in the prediction of low-pressure voxels, as demonstrated in Fig. 4.6.

of the simple semi-analytical models we use as a guess for the target. It should be emphasized that these semi-analytical models are not necessarily the best we could possibly construct; however, they should serve as useful reference points.

4.5.1 Electron pressure

The network was trained with electron pressure as target on the mentioned zoom-in simulations. The training set is relatively small (181 independent simulation boxes with enhanced resolution centered around massive halos). We observe the disadvantage of such a small training set in that the network starts overfitting after 176 epochs, when we halt training. Mixing samples from the TNG300 volume with the zoom-in simulations to enlarge the training set does not seem to yield improved performance in our experiments, indicating that the network updates are driven by the high-density regions in the extremely massive clusters.

In Fig. 4.5, we plot the electron pressure power spectrum; here, and elsewhere, computed using `nbodykit`¹¹ [222]. We plot two versions of the B12 semi-analytical model, the lower curve being the original output and the upper curve having the electron pressure values rescaled by a constant factor to make for a fairer comparison.

¹¹<https://github.com/bccp/nbodykit>

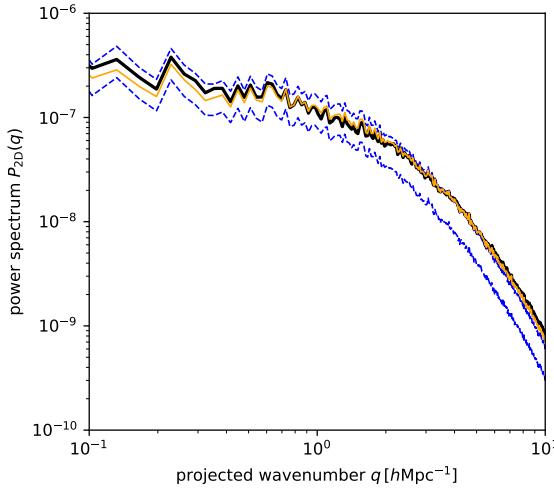


Figure 4.6: Electron pressure projected power spectrum, illustrating that upon projection along one axis the small electron pressure values, which the network is unable to recover correctly, are of lesser importance, compared to the 3-dimensional power spectrum in Fig. 4.5.

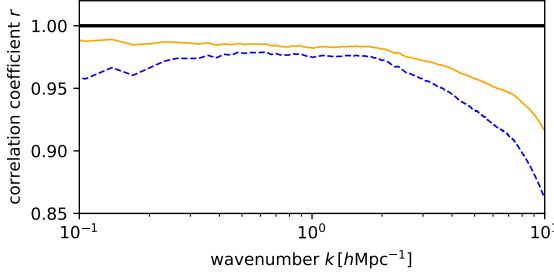


Figure 4.7: Electron pressure cross-correlation coefficient between network prediction/B12 semi-analytical model output and the target simulation field. Note that this measure is unaffected by the rescaling of the B12 semi-analytical model employed in Fig. 4.5.

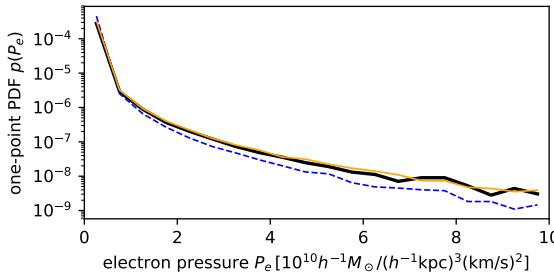


Figure 4.8: Electron pressure one-point PDF. The network prediction matches the target simulation very well. We only plot the original B12 semi-analytical model, not the rescaled version shown in Fig. 4.5.

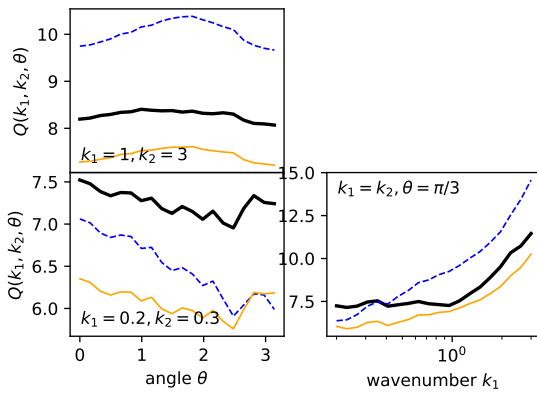


Figure 4.9: Some (reduced) bispectra for the electron pressure. Wavenumbers are given in units of $h\text{Mpc}^{-1}$. *Upper left:* small scales with varying triangle shapes, *lower left:* large scales with varying triangle shapes, *lower right:* varying scales with fixed equilateral triangle shape.

In terms of the power spectrum, the network’s prediction matches the TNG300 target result better than the B12 semi-analytical model on all scales and shows very good agreement with the target on small scales ($k \gtrsim 1 \text{ hMpc}^{-1}$). It should be emphasized, however, that B12 was calibrated on a different simulation.

The deficit in power on larger scales ($k \lesssim 1 \text{ hMpc}^{-1}$) is consistent with the zero-order approximation that the network neglects to predict electron pressures below a certain threshold (c.f. Fig. 4.3). We point out that if this explanation is correct, the lack of power is not particularly worrisome since upon projection along the line of sight (to obtain the Compton- y observable Eq. 4.1) the low pressure values will be further diminished in their contribution to any relevant summary statistics. In order to test this interpretation, we project the electron pressure along one cartesian axis and compute the power spectrum of the resulting two-dimensional field; this is plotted in Fig. 4.6. We observe that upon projection the agreement between the network prediction and the target improves, consistent with our explanation for the lack of power. It is quite conceivable that if one were to construct a complete tSZ map, i.e. project up to redshift 1100, the discrepancy would entirely disappear.

While the power spectrum measures the amplitudes of the different modes, the cross-correlation coefficient between two fields A and B

$$r_{AB}(k) \equiv \frac{P_{AB}(k)}{\sqrt{P_{AA}(k)P_{BB}(k)}} \quad (4.14)$$

is a measure of the correlation between their phases. We plot the correlation coefficient with respect to the target electron pressure in Fig. 4.7 (a perfect prediction would have unit cross-correlation with the target field). Again, the network matches the reference simulation better than the semi-analytical model and maintains a correlation of 98 % up to $k < 2 \text{ hMpc}^{-1}$.

While power spectrum and correlation coefficient are useful summary statistics, non-Gaussian statistics are of great importance for non-linear fields such as the tSZ. In Fig. 4.8, we plot the one-point PDF; again we observe very good agreement between the network prediction and the target, while the semi-analytical model appears to lack high-pressure voxels. This is a strong indication that the simulation details (in particular the amount of feedback) are sufficiently different between the IllustrisTNG simulations and the simulations the B12 semi-analytical model has been calibrated on that a direct comparison between the two is not particularly meaningful. Note that on linear P_e -scale the discrepancies at low values, which explain large-scale differences in the power spectrum, are not visible.

Another non-Gaussian statistic is the bispectrum, which we plot in Fig. 4.9 for several different triangle configurations; the computation was performed using `Pylians`¹² [223]. The network shows agreement with the target at the 20 % level. It should be pointed out that we did not use the bispectrum during cross-validation; one could imagine a training procedure where it is taken into consideration as well if more accurate predictions are needed. Another alternative would be to use a more general loss function if bispectrum predictions are desired. The large-scale bispectrum (lower left corner of Fig. 4.9) is the only summary statistics in this section for which the B12 semi-analytical model matches the simulation better than the network's prediction.

In order to gain more intuition, we present images of the electron pressure in several different halos in Fig. 4.10. The first three rows are taken from the TNG300 box; for the ones labeled “massive halo” and “typical halo” (with total halo masses of $M_{200c} = 8.8 \times 10^{14}$ and $2.6 \times 10^{14} h^{-1} M_\odot$, respectively); we observe very good agreement between network prediction and target. On the other hand, the network predicts too high electron pressure values for the “halo pair”, possibly it is unable to take the long-range interaction between the two halos into account, the number of

¹²<https://github.com/franciscovillaescusa/Pylians>

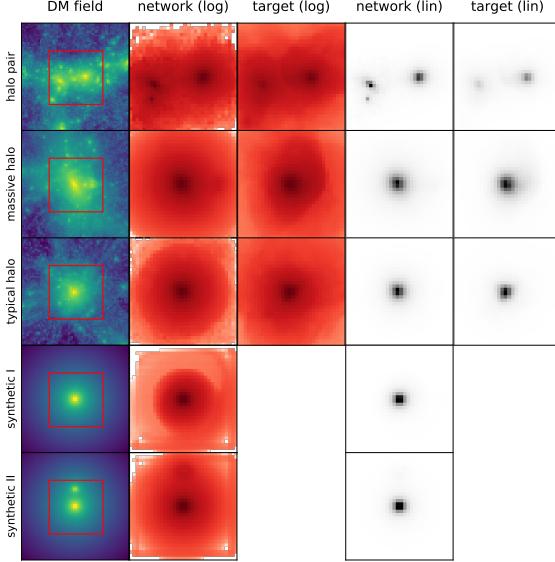


Figure 4.10: Some individual halos. We plot the matter density and electron pressure (both network prediction and simulation target), projected across 6.4 and $3.2 h^{-1}\text{Mpc}$ respectively. For the electron pressure, both logarithmic and linear scales are provided. The red boxes in the left column indicate the extent of the target boxes in the dark matter boxes (c.f. the discussion of padding in Appendix C.2). The first three rows are taken from the TNG300 box, while the last two are synthetic pure NFW dark matter halos for which we do not have a hydrodynamical simulation target available. Note that the color scales are different in each row in order to make interpretation easier.

such situations in the training set being very small. The last two rows are synthetic pure NFW dark matter profiles. For these we choose a halo mass of the primary object as $M_{200c} = 10^{14} h^{-1} M_\odot$; the last row also contains a smaller halo with $2 \times 10^{13} h^{-1} M_\odot$. Given these masses (and redshift $z = 0$), we use the concentration-mass relation from [224] and the density profile from [225] to create the “synthetic” input data. We observe that the network has learned spherical symmetry quite well (the small deviations near the edges of the box are completely irrelevant for the loss function, as evidenced by the plots in linear color scale).

The network performance for electron pressure is remarkable given that it was trained on the TNG-Cluster sample, which is unusually small (for U-Net standards) and highly biased, and which furthermore is expected to show some small-scale differences to the test set of TNG300 halos. We expect these small-scale differences because we do not have gravity-only zoom-in simulations available, so that the input data will be slightly different due to feedback and baryon-baryon interactions. The fact that these small-scale differences seem to be of minor importance indicates that the network does not assign much importance to features at the voxel scale; this fact gives us confidence in generalizability. The good performance despite strong biasing

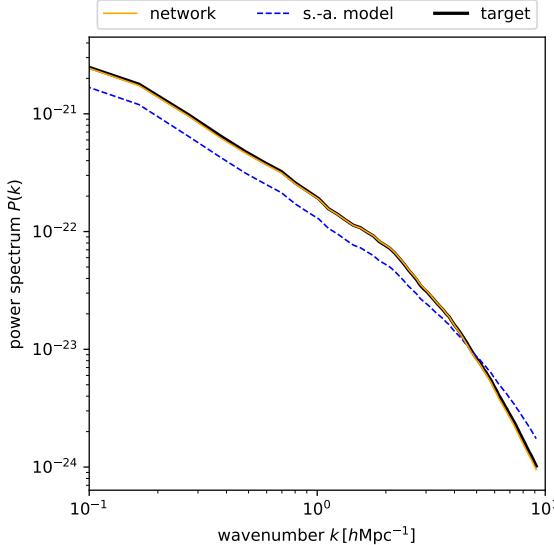


Figure 4.11: Electron density power spectrum. The network achieves excellent agreement with the target simulation, which is an indication that electron density is an easier target than both pressure and momentum density.

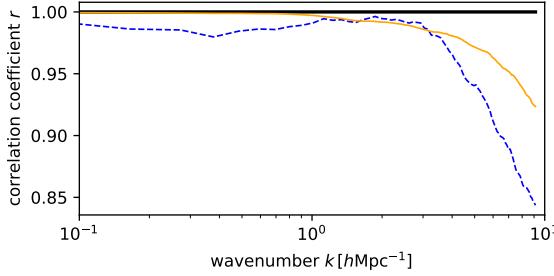


Figure 4.12: Electron density cross-correlation coefficient with the target simulation.

of the training set is interesting as well: we have mentioned that most of the halos in the training set are more massive than the most massive halos in the testing set. We interpret the fact that our procedure still works as an indication that the mapping between dark matter and electron gas may be dominated by sub-halo-scale features.

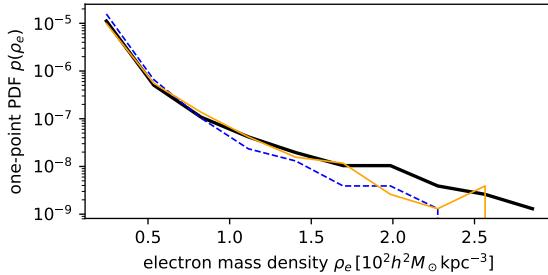


Figure 4.13: Electron density one-point PDF. Note that the test box is only 10 % the size used in Fig. 4.8, which explains the larger scatter in the tail.

4.5.2 Electron density

As mentioned in Section 4.4.1, learning the mapping from matter to electron density does not require such a strongly biased selection of the training samples as it did for the electron pressure. Thus, we can use sub-volumes of the TNG300 box as training samples, giving us a substantially larger training set. In this case, we do not observe any problems with overfitting. Since we expect the network to utilize quite similar features in its prediction of both electron pressure and density, we use the same network architecture and start the training for electron density with the best trained version for the electron pressure. Confirming our expectation, the network quickly (within about 20 epochs) adapts to the new target and converges slowly afterwards. Since no overfitting occurs, we train up to 210 epochs, when the validation loss shows no significant improvement anymore.

In Fig. 4.11, we plot the electron density power spectrum, which shows remarkable agreement between network prediction and target. Similarly, the cross-correlation coefficient with the target simulation, plotted in Fig. 4.12, exceeds 98 % up to $k < 4 \text{ } h\text{Mpc}^{-1}$. Finally, the one-point PDF, plotted in Fig. 4.13, also shows good agreement. The relatively noisy behaviour in the high-density tail (compared to Fig. 4.8) occurs because the volume of the testing box is only 10 % of the volume in the electron pressure testing set.

In summary, the electron density is an easier target compared to the electron pressure. Due to the better behaved distribution of target values, we are able to use a much larger training set¹³, preventing any problems with overfitting and leading to very good performance on the testing set. The fact that, apart from the change in training set, no modification to the network architecture and training procedure was

¹³While we have 181 primary objects in the TNG-Cluster zoom-in simulations available, there are 1416 halos in the TNG300 with masses larger than the cut-offs in Eqs. (4.9), (4.10).

necessary, indicates that the techniques to confront the sparsity problem, explained in Section 4.4, generalize well to different datasets¹⁴.

4.5.3 Electron momentum density

As in the case of the electron density, we find that we do not require the additional zoom-in simulations to learn the mapping to the momentum density; thus, the training set is identical to the one used in the previous section (the sample selection function is slightly different though, c.f. Section 4.4.1).

The electron momentum density is somewhat special in that \mathbf{p}_e is a three-component vector (while P_e and ρ_e are positive scalars). As mentioned in Section 4.3.1, one input channel is the dark matter density, and we include information on the dark matter momentum density in the input as well. Then we are left with three possible input-output configurations:

1. three directions as input, three directions as output;
2. three directions as input, one direction as output;
3. one direction as input, one direction as output.

The shape of the semi-analytical model would be chosen equal to the desired output shape. Empirically, we find options 1 and 2 not to work well with network architectures analogous to the one described in Section 4.3.3. Presumably, the network finds it difficult to learn that it has to establish three separate paths, with small coupling between them. This problem could possibly be solved by modifying the network architecture more fundamentally, however, in the interest of consistency, we choose not to do this and opt for the input-output configuration 3. This necessitates only minor modifications to the network architecture, as described in the caption to Fig. 4.2. It should be pointed out that this choice has some drawbacks: First, it is probably

¹⁴This is not obvious, since many hyperparameters were tuned during extensive trials on the electron pressure, while the only tuning necessary for the electron density was the biasing of the sample selection function, Eq. 4.9.

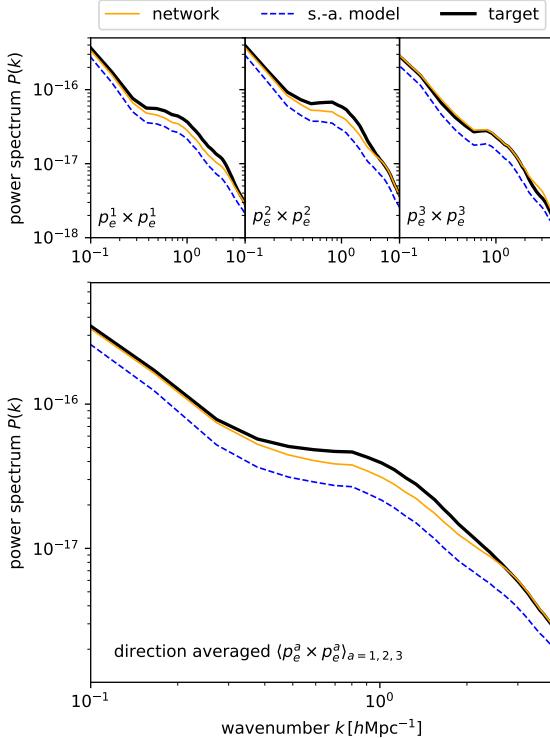


Figure 4.14: Electron momentum density power spectrum. While the upper panels show the power spectra of the momentum components in three cartesian directions, the lower panel is the average. Although the network performs much better than the semi-analytical model, some discrepancies ($\sim 20\%$) are clearly visible. They are likely related to a combination of overfitting and a sub-optimal network architecture.

inefficient, since many features can be assumed to be shared between the three different spatial directions. Second, we could imagine situations in which the dark matter velocity in a particular voxel satisfies inequalities like $v_{\text{DM}}^1 \ll v_{\text{DM}}^2$ (with the superscripts representing different cartesian coordinate directions) and we would like to predict p_e^2 . In such a situation the network would be unable to incorporate small rotations relating the dark matter and electron velocity directions (which would make v_e^2 drastically different from v_{DM}^2), leading to inaccurate predictions of p_e^2 .

Besides the described modifications to the network architecture, we use the tuned sample selection function of Eq. (4.10), and find that increasing the noise to 20 % (c.f. Section 4.3.4) yields better performance. We observe slight overfitting and halt training after 188 epochs.

In Fig. 4.14, we plot electron momentum density power spectra. The three plots in the top row represent the three directions, while the larger plot in the bottom row is simply the average of these. Of course, taking this average is not a particularly

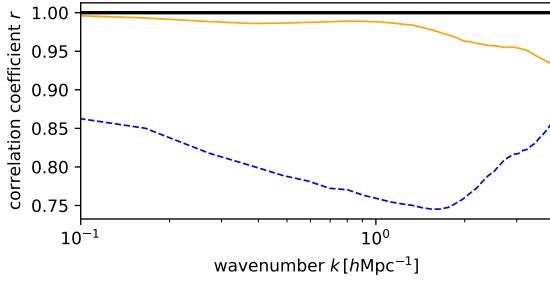


Figure 4.15: Electron momentum density cross-correlation coefficient with the target simulation, averaged over the three cartesian directions. The scatter between different directions is small.

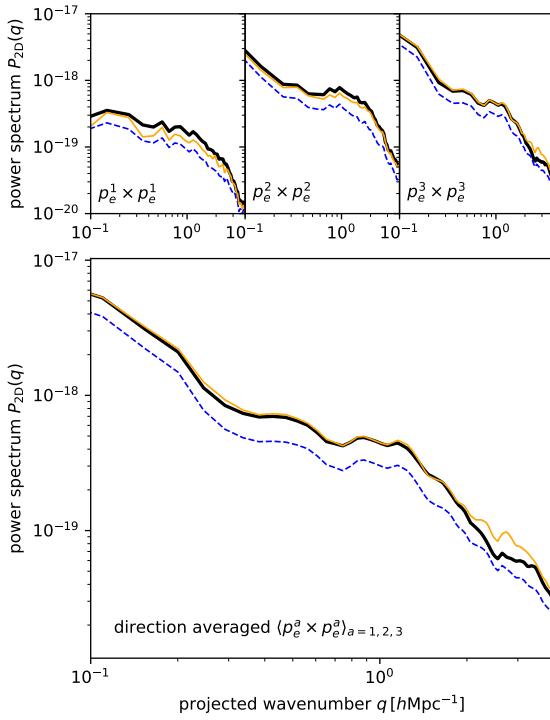


Figure 4.16: Electron momentum density projected power spectrum. The format is the same as in Fig. 4.14.

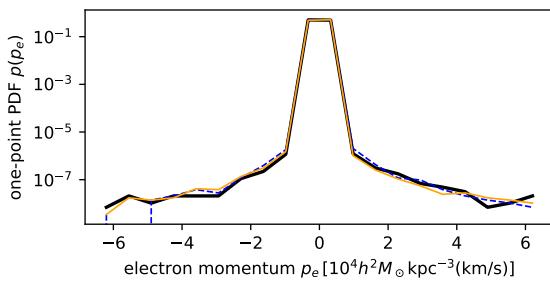


Figure 4.17: Electron momentum density one-point PDF. Again, all three cartesian directions are included.

meaningful operation, given that the momentum components are correlated, but it should serve as a less noisy representation of the network performance. We see that although the network outperforms the simple semi-analytical model, it shows a power deficit of up to 20 % in the intermediate k -range. Interestingly, we actually observed a slight power excess when evaluating the network on the validation set, indicating that we overfit on the validation set. As a useful proxy for the actual kSZ-observable, we plot the projected power spectra in Fig. 4.16 (projections are along the individual momentum directions, which is the physically relevant case). We observe a large variability in the power spectra for the three different directions (top row), making the averaged version less meaningful. It can be seen that the power deficit observed in the 3D power spectra is still present in the projected versions, although upon averaging it is less prominent.

The correlation coefficient shown in Fig. 4.15 is an average over the three directions, we did not observe significant variability between the directions. We recover correlations exceeding 96 % up to $k < 2 \text{ } h\text{Mpc}^{-1}$, indicating that the predicted momentum field has somewhat lower fidelity than in the cases of electron pressure and density. However, we are comparing correlations at an excellent level here.

It should be noted that the semi-analytical model performs significantly worse than in the other two cases (compare Figs. 4.7, 4.12). Since the network builds on the semi-analytical model, worse performance should be expected. From the one-point PDF in Fig. 4.17 we see that network and semi-analytical model perform about equally well on this summary statistics. One conclusion to be drawn from this is that the deficit in power exhibited by the semi-analytical model cannot be solved by simply rescaling it.

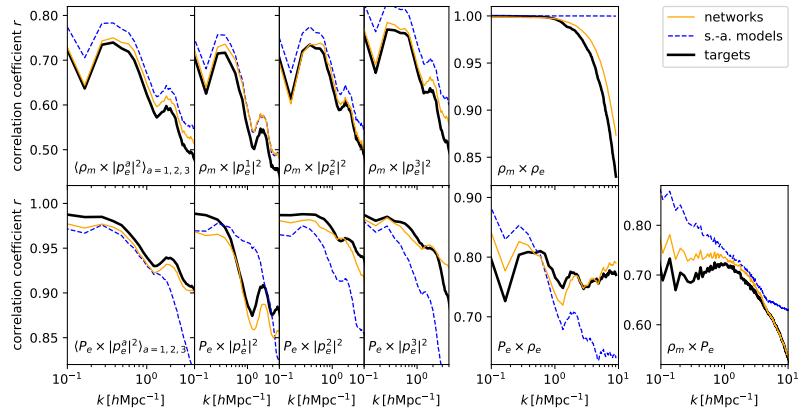


Figure 4.18: Cross-correlation coefficients between different fields. We observe relatively good agreement between network prediction and target simulation for a range of observables and scales, but discrepancies for others. The networks consistently outperform the semi-analytical models, but it is likely that the network prediction quality depends significantly on the semi-analytical model’s quality.

4.5.4 Cross-correlations between different fields

As the last part of the section on results, we present cross-correlations between different fields evaluated on the testing set. The purpose of this is twofold. First, correlating different cosmological observables can be less susceptible to systematic effects, which makes them useful summary statistics [e.g., for weak lensing cross tSZ, 226–228]. Second, these cross-correlation functions were not considered during training, making them a “double blind” way of evaluating the network performance (unseen data and new types of summary statistics) and therefore they represent a useful diagnostic.

Fig. 4.18 shows the 3-dimensional cross-correlations. Note that the ρ_m -field is the same for all correlations including it, since it comes from the gravity-only simulation.

- $\rho_m \times |\mathbf{p}_e|^2$ (element-wise squared): The top left panel in Fig. 4.18 is direction averaged (with the caveats discussed in Section 4.5.3), while the narrower panels to its right are for the individual momentum directions. This type of correlation function can be seen as a proxy for weak lensing \times kSZ². The network vastly

outperforms the semi-analytical model on larger scales ($k \lesssim 0.5 \text{ } h\text{Mpc}^{-1}$), while on smaller scales its performance is about a factor two better.

- $P_e \times |\mathbf{p}_e|^2$ (element-wise squared): The panels are arranged analogously to the previous case. This correlation function emulates tSZ \times kSZ² correlations. We observe that the semi-analytical models struggle to generate remotely correct predictions for this correlation function (although the average, perhaps by chance, happens to look reasonable for $k \lesssim 1 \text{ } h\text{Mpc}^{-1}$). The networks have problems with this correlation as well, presumably because it combines two predicted fields (while ρ_m in the previous case was correct by design).
- $\rho_m \times \rho_e$: This correlation is mostly a diagnostic and not related to any real-world observable. As expected from the design of the semi-analytical model, Eq. (4.7), the semi-analytical model’s prediction for ρ_e is perfectly correlated with the matter field. The network’s prediction is closer to the target correlation function.
- $P_e \times \rho_e$: Again, this plot is a diagnostic. As observed before, correlating two generated fields proves to be a challenge for the semi-analytical models. The networks’ prediction is about a factor two better than the semi-analytical models’ for $k \lesssim 1.5 \text{ } h\text{Mpc}^{-1}$ and much better on smaller scales. We observe the trend already remarked in Section 4.5.1 that the network prediction for P_e is less accurate on large scales. However, unlike the $P_e \times P_e$ auto-power, projecting the field does not seem to substantially improve the agreement between network prediction and simulation target.
- $\rho_m \times P_e$: This correlation is a proxy for weak lensing \times tSZ statistics. Again, the network struggles on large scales with a correlation function involving P_e , while the small-scale predictions are excellent.

We said at the beginning of this section that considering the cross-correlations has two purposes. With respect to real-world data analyses, we observe high-fidelity predictions for a subset of the correlations and not necessarily in the whole range

of wavenumbers considered. However, we point to the fact that, as observed before, projecting the fields can help to bring the networks' predictions closer to the truth.

With respect to the diagnostic purpose of these plots, we first note that the discrepancies observed here are roughly in line with those seen in different Gaussian summary statistics before, indicating that we did not strongly overfit on the type of summary statistics. A final comment is that we find the discrepancies between network predictions and truth to be mostly in the direction of the discrepancies between semi-analytical models and truth. This is an indication that the network performance depends substantially on the semi-analytical model's quality.

4.6 Conclusions

As part of an effort to quickly generate maps of non-linear observables, we have developed machine learning techniques that allow us to map the gravity-only matter field to three electron gas properties, namely pressure, density, and momentum density. This is the first step towards the generation of Sunyaev-Zel'dovich maps solely from gravity-only simulations, or even from high-redshift initial conditions.

In this work, we have focused on the three-dimensional fields at redshift $z = 0$. We have trained a deep convolutional neural network with a U-Net architecture using the output of cosmological gravity-only and full-physics hydrodynamical simulations from the IllustrisTNG project, specifically the TNG300 box and TNG-Cluster, a set of zoom-in cluster simulations. We have shown that the network outputs match the reference target simulation better than the predictions from simple semi-analytical models, according to a range of summary statistics. To achieve this, we have encountered the problem of sparse data sets, and explained a number of techniques that enabled us to solve it.

While we cannot, at present, exactly quantify the speed-up gained by our technique, it is clear that it will enable us to generate much larger sky areas than would be possible with computationally expensive hydrodynamical simulations. As point of reference, running the reference simulation used in this work, TNG300, took 34.9M CPU-hours, while evaluating the neural network on the same volume took 16 GPU-hours. In order to construct a light-cone, we would probably require evaluation on a few dozen redshift slices.

Some interesting findings made in the course of this work, and explained in more detail in previous sections, are: (1) the networks do not assign much weight to voxel-scale features; (2) the mapping between dark matter and electron pressure is dominated by sub-halo-scale features; (3) electron density is a much easier target than pressure, indicating that the degree of sparsity influences the network prediction quality.

Sub-optimal network performance in some cases (kSZ, some cross-correlations) can be explained by three factors: (1) scarcity of training data (leading to overfitting in some cases); (2) sub-optimal network architecture for the kSZ effect; (3) sub-optimal quality of the semi-analytical models. We believe that these problems are technical in nature and do not present a fundamental obstacle. For example, one could re-calibrate the B12 semi-analytical model for different simulations.

This work is only the first step in a longer program. While we only work at redshift $z = 0$, this is the prerequisite for the prediction of light cones. A further obvious and easy generalization would be to the relativistic tSZ effect [229–231]. Another extension would be to X-ray maps, for which excellent observational data will be available from ROSAT and eROSITA. Due to the signal’s dependence on the square of the electron density the sparsity problem is likely quite severe; the data augmentation with zoom-in simulations which we used for the tSZ effect will be essential in that case.

In terms of future work, a clear goal is the generalization to redshifts $z \neq 0$, which will ultimately enable us to generate maps of actual observables. There is a good chance that the dependence on redshift is a relatively simple function; this would imply that we would effectively enlarge the training data set and mitigate the overfitting problem. While we have worked with a single simulation (TNG300) so far, there is a need to compare networks trained on different simulations (one could also imagine mixing samples with different sub-grid physics, although the outcome of such a training strategy should be validated thoroughly). A possible route could be to use the CAMEL simulations [232], which contain state-of-the-art hydrodynamical simulations spanning thousands of different cosmologies and astrophysics models run with two different hydrodynamical solvers and subgrid implementations.

As mentioned in the Introduction, our approach does not capture stochasticity in the mapping from gravity-only matter to electron gas. The probabilistic nature stems from unresolved features that are coupled to the scales of interest, for example, from the stochasticity in the efficiency of AGN feedback. Possible extensions to this work which would take this effect into account could involve GANs [233] or Bayesian neural networks [234]. We note that with GANs it would be challenging to generate large enough volumes such that the two-halo term is captured correctly. A further difficulty in both approaches would be to create consistent predictions of several different observables, as is required for the cross-correlation functions. However, we believe that stochasticity does not present a fundamental obstacle to our approach, since it could accurately be modeled as additional white noise in any parameter inference.

Our work opens up the avenue towards several possible applications. Given the ability to quickly generate maps of the Sunyaev-Zel'dovich effects and other observables, we could create covariance matrices and perform parameter inference from summary statistics. In order to utilize the non-Gaussian information in maps, likelihood-free inference [235] could be a promising approach; for example applied to the tSZ one-point

PDF [236, 237]. Furthermore, we could imagine interpreting the network output to learn more about the effect of sub-grid astrophysical models.

Acknowledgements. We would like to thank J.C. Hill and E. Komatsu for useful suggestions. The training of the neural networks has been carried out on the Tiger supercomputer of Princeton University. We thank their system administrators for their help at several stages of this work. FVN acknowledges funding from the WFIRST program through NNG26PJ30C and NNN12AA01C. The Flatiron Institute is supported by the Simons Foundation. The IllustrisTNG simulations including TNG300 were undertaken with compute time awarded by the Gauss Centre for Supercomputing (GCS) under GCS Large-Scale Projects GCS-ILLU and GCS-DWAR on the Hazel Hen supercomputer at the High Performance Computing Center Stuttgart (HLRS). The TNG-Cluster simulation has been executed on Draco and Cobra machines of the Max Planck Computing and Data Facility (MPCDF) in Garching, Germany.

Chapter 5

Predicting the Thermal Sunyaev-Zel'dovich Field using Modular and Equivariant Set-Based Neural Networks

Abstract. Theoretical uncertainty limits our ability to extract cosmological information from baryonic fields such as the thermal Sunyaev-Zel'dovich (tSZ) effect. Being sourced by the electron pressure field, the tSZ effect depends on baryonic physics that is usually modeled by expensive hydrodynamic simulations. We train neural networks on the IllustrisTNG-300 cosmological simulation to predict the continuous electron pressure field in galaxy clusters from gravity-only simulations. Modeling clusters is challenging for neural networks as most of the gas pressure is concentrated in a handful of voxels and even the largest hydrodynamical simulations contain only a few hundred clusters that can be used for training. Instead of conventional convolutional neural net (CNN) architectures, we choose to employ a rotationally equivariant DeepSets architecture to operate directly on the set of dark matter particles. We argue that set-based architectures provide distinct advantages over CNNs. For example, we can enforce exact rotational and permutation equivariance, incorporate existing knowledge on the tSZ field, and work with sparse fields as are standard in cosmology. We compose our architecture with separate, physically meaningful modules, making it amenable to interpretation. For example, we can separately study the influence of local and cluster-scale environment, determine that cluster triaxiality has negligible impact, and train a module that corrects for mis-centering. Our model improves by 70 % on analytic profiles fit to the same simulation data. We argue that the electron pressure field, viewed as a function of a gravity-only simulation, has inherent stochasticity, and model this property through a

conditional-VAE extension to the network. This modification yields further improvement by 7%, it is limited by our small training set however. We envision that our method will prove useful in problems beyond the specific one considered here.¹²

5.1 Introduction

A pressing problem in cosmology is the accurate modeling of observables sourced or influenced by physics beyond gravity, in short, baryonic effects. Hydrodynamic simulations are the canonical forward model for such fields; however, their computational cost is too high for them to be a viable contender in generating the vast number of realizations necessary to sample distributions. Thus, an approach that has recently emerged is the use of neural networks to map cheaper gravity-only simulations to their full-physics counterparts. Not only does this idea enable a substantial speed-up in generating realizations, but it could also improve our physical understanding; to this aim interpretable models are required.

One example of a baryonic observable that we aim to model is the thermal Sunyaev-Zel'dovich (tSZ) effect [73, 74, 239, 75, 240]. The tSZ field is a secondary cosmic microwave background (CMB) anisotropy generated by inverse Compton scattering between electrons in the Universe's large scale structure and CMB photons. Its intensity at sky position \hat{n} , to leading order determined by the Compton- y parameter, is given by a line-of-sight integral over the electron pressure, $P_e(\vec{x})$:

$$y(\hat{n}) = \frac{\sigma_T}{m_e c^2} \int dl P_e(\hat{n}, l). \quad (5.1)$$

The problem that we tackle in this work is the prediction of the electron pressure $P_e(\vec{x})$ given a gravity-only simulation. Since this is a translationally equivariant spatial problem, the seemingly natural approach chosen for similar problems, e.g., by

¹We make our code publicly available at this URL. Data products, trained models, and hyperparameter SQL databases will be shared upon reasonable request.

²based on Ref. [238], joint work with M. Cranmer, W. Coulton, S. Ho, and D. N. Spergel.

[241–244, 156, 245, 246], is a convolutional neural net (CNN), taking as input the density field of a gravity-only simulation. However, in this work we argue that existing domain knowledge on $P_e(\vec{x})$ and similar fields renders the CNN approach inferior to a set-based architecture. In fact, electron pressure values high enough to affect observables are predominantly found in massive gravitationally collapsed structures, called clusters. A neural network should naturally take this property into account. From this point of view, translational invariance is in fact broken, negating the main advantage of CNNs.

Besides the heavy bias towards clusters, the electron pressure field exhibits a related feature that further inhibits the performance of CNNs. Even within a cluster the mass and electron pressure is highly concentrated towards the center. Thus, CNNs operating on the density field are highly inefficient in that they need to maintain a uniformly high spatial resolution in order to resolve the important small-scale details near the cluster center, while conversely this high resolution leads to a waste of resources in the low-density outer regions.

To a first approximation, clusters are described by their mass, M_{200} , and radius, R_{200} . Here, we employ the standard convention of mass within a spherical volume that contains an average density 200 times larger than the mean density of the Universe. These quantities determine a characteristic pressure scale: $P_{200} \propto M_{200}/R_{200}$. The electron pressure³ $P_e(\vec{r})$ is to leading order a spherically symmetric function, commonly approximated as a generalized Navarro-Frenk-White (GNFW) profile [225, 247, 147],

$$P_e(\vec{r}) \approx \text{GNFW}(|\vec{r}|; M_{200}, R_{200}), \quad (5.2)$$

which we will use as the benchmark (with the parameterization as chosen in [147], fitted to our data). There is an inherent random element in the electron pressure field if viewed as a function of a gravity-only simulation’s snapshot at a given time. The

³We use \vec{r} for coordinates relative to a cluster’s position.

reason is that chaos washes out some of the history; in particular the time-integrated activity of the active galactic nuclei (AGNs) is difficult to infer.

We propose to learn a probabilistic mapping directly from the simulation representation, i.e., from a set of dark matter particles with associated positions \vec{q}_i and velocities \vec{v}_i . For a cluster α our most general model can be written as

$$\hat{P}_e(\vec{r}) = F(\{(\vec{q}_i^{(\alpha)}, \vec{v}_i^{(\alpha)})\}_{i \in \alpha}; \{(\vec{q}_i^{(\vec{r})}, \vec{v}_i^{(\vec{r})})\}_{|\vec{q}_i - \vec{r}| < R}; \mathbf{s}_\alpha, \mathbf{e}_\alpha; \mathbf{a}; \vec{r}), \quad (5.3)$$

where \mathbf{s}_α are scalar properties describing the cluster, \mathbf{e}_α are unit vector properties, $\mathbf{a} \sim \mathcal{N}(\mathbf{0}, \mathbb{I})$ is drawn from a standard normal, and we distinguish between feature tuples and $\text{SO}(3)$ vectors using the given notation. The first argument to F is the set of dark matter particles comprising the cluster, positions and velocities are evaluated relative to the cluster position and bulk motion respectively. Conversely, the second argument is the set of particles in the vicinity of the target position \vec{r} , where the positions are relative to \vec{r} and the velocities relative to the local bulk motion; R is a hyperparameter. Note that operating on particle sets naturally solves the resolution problem mentioned earlier.

DeepSets [248] are a class of architectures that naturally operate on such sets, part of a broader direction in deep learning known as “geometric deep learning” (see for example [249–252] and citations therein). Geometric deep learning has previously seen success in cosmology in works such as [253–256]. Given a tuple of scalars \mathbf{f}_i associated with the i th dark matter particle, a DeepSet first computes another tuple \mathbf{g}_i using a multi-layer perceptron (MLP). Then a pooling operation (in our case the mean) over the i -direction produces a feature tuple that is invariant under the ordering of the input particles. We denote such an architecture as a scalar DeepSet. We construct the input features \mathbf{f}_i so as to make its elements $\text{SO}(3)$ scalars [257]. This can be achieved by using properties such as $|\vec{q}_i|$, $|\vec{v}_i|$, and contractions between \vec{q}_i , \vec{v}_i and the elements

of \mathbf{e}_α . A simple extension multiplies the \mathbf{g}_i with the \vec{q}_i before pooling, thus leading to an output feature tuple in which each element is an SO(3) vector. We denote such an architecture as a vector DeepSet. It is easy to see that the described vector DeepSet is rotationally equivariant, since its output is a linear combination of SO(3) vectors with SO(3) scalar coefficients. Thus we obtain a rotationally equivariant class of architectures operating directly on the particle representation instead of gridded fields.

It should be noted that DeepSets exhibit a further feature that turns out to be useful in applications similar to the one presented here. Namely, given an appropriate pooling operation, they are to leading order invariant under the number of input particles. Thus, training can be performed efficiently on small input sets while inference is then possible, with typically higher accuracy, on larger sets. This ‘sparsification’ of input sets during training is also a form of regularization; in fact, we find it crucial in mitigating overfitting.

5.2 Architecture

Fig. 5.1 schematically illustrates the various architecture components. We emphasize that most modules can be trained and evaluated independently. This modular design makes the architecture amenable to interpretation. At the end of Sec. 5.4 we will briefly mention several modules we have experimentally added to the architecture. At various points the cluster-scale properties \mathbf{s}_α , \mathbf{e}_α are passed, which we omit for conciseness.

The function f produces the final output $\hat{P}_e(\vec{r})$ using two components, namely a (modified) GNFW prediction and the output of the *Aggregator* MLP. By construction, the GNFW prediction is a single scalar g , while we let the *Aggregator* output two

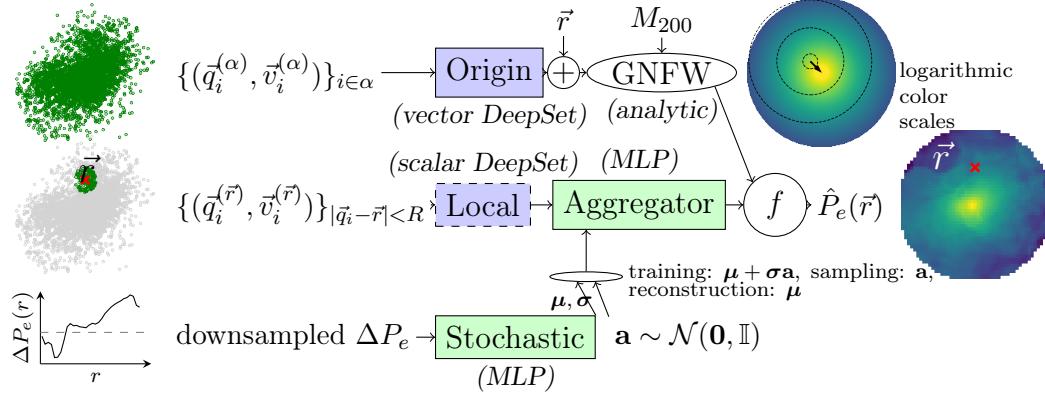


Figure 5.1: Schematic illustration of the architecture used in this work. Our task is to predict the electron pressure P_e at positions \vec{r} , using phase space coordinates \vec{q}, \vec{v} of the dark matter particles. The spherically symmetric analytic GNFW profile generally provides a good approximation for the target electron pressure field and also serves as the benchmark in this work. The first correction we introduce is for mis-centering. This correction to the point where the GNFW profile is centered is computed by the *Origin* module which is a vector DeepSet operating on all particles in the halo. Second, we introduce local corrections to the GNFW profile using the *Local* module which is a scalar DeepSet operating on all particles within a certain radius R from the target position. Third, stochasticity is modeled using a conditional-VAE design, in which the *Stochastic* module is an MLP acting on a downsampled version of the residuals. The *Aggregator* MLP and the function f serve to combine and compress their inputs. Individual modules can be independently removed from and added to the architecture.

scalars, a_1, a_2 . The function f is chosen as:

$$\hat{P}_e(\vec{r}) = f(g, a_1, a_2) = \text{ReLU}[g \text{ReLU}(1 + a_1) + k \sinh a_2], \quad (5.4)$$

where k is a learnable parameter and $\text{ReLU}(x) \equiv \max(0, x)$ is the rectifying linear unit. The outer ReLU is chosen because the output is known to be positive, and the inner ReLU because we believe it unphysical for the GNFW prediction to contribute with inverted sign. The sinh function provides flexibility in that it can act both as a linear and an exponential, depending on the hyperparameter k , a similar parameterization has already been found useful in [156]. In some of our experiments we prune the network such that the entire *Aggregator* disappears; in such cases f simply returns the GNFW prediction g .

The GNFW model takes as input the target radial position $|\vec{r}|$, which is corrected for mis-centering by the *Origin* module (the cluster finder estimates cluster positions that are not necessarily best to center the GNFW profile at). The *Origin* module is a vector DeepSet operating on the input features $|\vec{q}|$, $|\vec{v}|$, $\hat{q} \cdot \mathbf{e}_\alpha$, $\hat{v} \cdot \mathbf{e}_\alpha$, and \mathbf{s}_α for particles within $2.5 R_{200}$. The $|\vec{q}|$ and $|\vec{v}|$ are normalized by R_{200} and V_{200} respectively, and afterwards transformed to zero mean, unit variance over the training set. The *Origin* module outputs two vectors \vec{o}_1, \vec{o}_2 ; the final shift in the cluster center is then computed as

$$\Delta \vec{o} = \frac{R_{200}}{2} \tanh[(X_{\text{off}}/R_{200})\vec{o}_1 + \vec{o}_2], \quad (5.5)$$

with X_{off} as defined in the following section. It appears reasonable that in many cases the desired shift in origin will correlate well with a spatial measure of relaxedness (such as X_{off}), which motivates the \vec{o}_1 term. However, in some cases this assumption may be false (e.g., X_{off} happens to be very small), in which case we give the network the freedom to disregard our physical intuition via the separate \vec{o}_2 vector. The tanh acts element-wise and stabilizes training. In the fully trained network, $\Delta \vec{o}/R_{200}$ is generally small, so the tanh reduces to a linear (in which case $\Delta \vec{o}$ transforms as a vector to a good approximation). However, due to the non-linear effects $\Delta \vec{o}$ has on other components of the network it is useful to bound its magnitude to a reasonable range so that the training process does not become stuck in pathological configurations.

The *Aggregator* MLP combines multiple inputs.

The *Local* module produces scalar features from the set of dark matter particles in the vicinity of the target position, where the cutoff R is a hyperparameter. For each particle we pass the features $|\vec{q}|$, $|\vec{v}|$, $\hat{q} \cdot \mathbf{e}_\alpha$, $\hat{v} \cdot \mathbf{e}_\alpha$, and $\vec{q} \cdot \vec{v}$, where \vec{q} and \vec{v} are with respect to the target position and local bulk velocity respectively and their magnitudes normalized to zero mean, unit variance over the training set. After passing the local particles through the DeepSet, we concatenate the resulting tuple with the number of particles within R to set the scale. If other cluster-scale information, besides the

Local module, is passed to f , we also pass information on the target position \vec{r} to the *Aggregator*. This is achieved by passing the scalars $|\vec{r}|$ and $|\hat{r} \cdot \mathbf{e}_\alpha|$, with $|\vec{r}|$ normalized by R_{200} , as well as the \mathbf{s}_α . The use of the *Local* module is a form of expanding our training set (one could imagine compressing the set of all particles comprising the cluster into some code which is then evaluated at different positions; in this case, the training set would be very small however).

The other input to the *Aggregator* models the probabilistic nature of the mapping, through a conditional VAE [258–261] architecture. The *Stochastic* module is the standard VAE encoder, taking as input the residuals of the electron pressure field with respect to a deterministic model. Since the simulation we are using implements AGN feedback in a spherically symmetric fashion, we average these residuals in spherical shells around the cluster position (another instance of domain knowledge). These spherical shells are chosen with respect to the position computed by the Rockstar halo finder (see the following section), i.e., the *Origin* module has no influence on how these shells are chosen. These shells are chosen as 32 linearly spaced radial intervals out to $2R_{200}$. We perform principal component analysis on the residuals over the training set and find that the most important eigenvector represents transfer of electron pressure between inner and outer parts of the cluster, bolstering our intuition that the stochasticity is driven by AGN feedback. We will find in Sec. 5.4 that this interpretation is perhaps too simplistic, however. Given our expectation that the unresolved AGN activity drives stochasticity, we choose the VAE code to be a single number; such a small latent space dimension is also useful in regularizing the model in view of the small training set. We have explicitly confirmed that doubling the latent space dimension leads to worse results.

On a fundamental level, our architecture composes many MLPs. While we allow for some degree of heterogeneity in their structure, we generally use LeakyReLU activation functions, layer normalization [262], depths of $3 \cdots 4$ hidden layers, and

~ 196 hidden neurons per layer. We find that, not surprisingly, the *Aggregator* is most susceptible to overfitting. This motivates us to introduce dropout [263] there, with rates of $\sim 5\%$ found optimal in the hidden layers while the first layer requires much higher rates of $\sim 30\%$.

5.3 Data and training

We use the IllustrisTNG 300-1 simulation [133–138] for training and testing. This simulation provides a gravity-only and a full-physics run with the same initial conditions. In this work, we restrict ourselves to the present-day (redshift $z = 0$) snapshot; a generalization to earlier times is naturally possible by passing z to the various MLPs in the architecture.

We use the state-of-the-art cluster finder code Rockstar [264, 265] to identify clusters with masses $M_{200} > 5 \times 10^{13} M_\odot/h$ in the gravity-only snapshot.⁴ The resulting 463 clusters are randomly assigned to training (70 %), validation (20 %), and testing (10 %) sets. They have radii R_{200} ranging from 600 to 1600 kpc/h and contain between 1.5 and 47 million dark matter particles within $2.5R_{200}$.

The cluster-scale scalars \mathbf{s}_α are chosen as:

- logarithmic mass, $\log M_{200}$,
- position offset from center of mass, X_{off} ,
- velocity offset from bulk motion, V_{off} ,
- same as X_{off} , but with respect to the center of mass of *all* particles within $2.5 R_{200}$ (instead of only the ones identified by Rockstar as being bound),
- magnitude of angular momentum, $|\vec{J}|$,
- eigenvalues of mass inertia tensor, $I_a^{(m)}$ ($a = 1, 2, 3$),

⁴The reason for this choice of mass cutoff is that in the IllustrisTNG astrophysics model at lower masses the gas physics changes qualitatively as AGN feedback is more effective in driving gas out of the cluster.

- eigenvalues of velocity inertia tensor, $I_a^{(v)}$ ($a = 1, 2, 3$) (sometimes called the velocity dispersion tensor).

The cluster-scale vectors \mathbf{e}_α are chosen as:

- angular momentum, \vec{J} ,
- position offset, \vec{X}_{off} ,
- eigenvectors of mass inertia tensor, $\vec{I}_a^{(m)}$ ($a = 1, 2, 3$),
- eigenvectors of velocity inertia tensor, $\vec{I}_a^{(v)}$ ($a = 1, 2, 3$).

It is important to ensure the correct behaviour of these quantities under SO(3) transformations. Thus, we order eigenvalues, together with the corresponding eigenvectors, according to magnitude. The orientation of the eigenvectors needs also to be fixed. This can be accomplished using any pseudo-vector, in this case we choose orientation such that the contraction of any eigenvector with the angular momentum \vec{J} is positive.

All vectorial properties \mathbf{e}_α are normalized such that contractions with isotropically distributed unit vectors have unit variance. The scalar properties, except for $\log M_{200}$, are normalized by the appropriate self-similar scales; using M_{200} and R_{200} , any quantity can be nondimensionalized. Thus, we reduce all inputs to geometric ones apart from $\log M_{200}$ which sets the scale. After the self-similar normalization, we apply the affine transformation to the \mathbf{s}_α that yields zero mean and unit variance over the training set.

Since our training set is relatively small, we find it crucial to add noise to the cluster-scale properties. The non-uniform distributions of some of the scalars \mathbf{s}_α motivate us to use the following procedure. For each feature in \mathbf{s} , we sort the values in the training set and compute the nearest-neighbour differences Δ_+, Δ_- . During training, we generate noise according to

$$\delta = Nr\Delta_{\text{sgn}(r)}, \quad r \sim \mathcal{N}(0, 1). \quad (5.6)$$

Here, N is a hyperparameter for which we find values $\mathcal{O}(10)$ to work best. Owing to its exceptional role as setting the overall scale, we allow for a different value of N for the $\log M_{200}$ feature, finding that somewhat lower noise levels are optimal. Note that this noise prescription does not actually preserve the mean, but empirically we find it to work better than ‘patching’ two one-sided Gaussians. For the vector properties, we experimented with rotations by angles drawn from a normal. We find mild evidence that standard deviations of ~ 2 deg are preferred over noiseless vectors.

We produce electron pressure fields from the full-physics simulation using Voxelize [266], with a voxel sidelength of $5R_{200}/64$. This resolution is sufficient for current and near-future tSZ measurements; we find that higher resolutions lead to more unstable training, presumably due to large local outliers. The choice to keep the resolution in units of R_{200} is motivated by the general geometric spirit of this work, a production-quality model would most likely choose a constant resolution instead.

Our reconstruction loss for a given cluster is given by

$$\mathcal{L}_{\text{recon}} = \left\langle \left(\frac{P_e(\vec{r}) - \hat{P}_e(\vec{r})}{P_{200}} \right)^2 \right\rangle_{|\vec{r}| < 2R_{200}}, \quad (5.7)$$

where the normalization with P_{200} mitigates our dearth of clusters at the high-mass end.⁵ The target positions \vec{r} are randomly sampled during training for efficiency, while testing is of course performed on all available voxels.

Training is performed using the Adam optimizer [267] and a one-cycle learning rate schedule [268]. Each epoch iterates once through all clusters in the training set, randomly choosing 256 target positions in each cluster and assembling 4 clusters into each mini-batch. We typically train for 100 epochs, in some cases 200. We find it necessary to choose different learning rates for the individual modules. For modules that have access to cluster-scale information (*Aggregator*, *Origin*), we also apply weight

⁵For practical applications the scaling with P_{200} should be omitted, in which case a somewhat larger training set will likely be required.

decay. We find it beneficial to apply gradient clipping (typically at $\mathcal{O}(1)$) in order to stabilize training.

It turns out to be optimal to pass $\mathcal{O}(10^3)$ randomly sampled dark matter particles to the *Origin* module during training, since such a sparse sampling reduces overfitting. We typically pass $\mathcal{O}(10^2)$ dark matter particles to the *Local* module, this is purely for efficiency. In many cases there are fewer dark matter particles than this number in the local vicinity; then we simply sample some particles multiple times. During testing we increase these numbers typically by a factor $\mathcal{O}(10)$ and observe some benefit from the finer sampling. Further increase of the number of dark matter particles passed does not increase the prediction quality.

For hyperparameter searches we use the Optuna package [144], solving the problem

$$\theta_{\text{opt}} = \operatorname{argmin} \mathcal{L}_{\text{opt}}(\theta) \text{ with } \mathcal{L}_{\text{opt}}(\theta) \equiv \text{median}(\mathcal{L}_{\text{recon}}[\text{network}_\theta]/\mathcal{L}_{\text{recon}}[\text{GNFW benchmark}]), \quad (5.8)$$

where the median is over the validation set at the end of training. During training runs on architectures in which the stochastic module is included, we take as the training loss the sum of reconstruction loss and negative KL divergence of the VAE code with respect to a standard normal, the latter multiplied with a scaling which we anneal from zero to some hyperparameter over the course of training. For such architectures, we perform multi-objective optimization on both \mathcal{L}_{opt} and the mean of the KL divergence over the validation set. We then produce figures similar to Fig. 5.2 for multiple models in the Pareto frontier, considering the validation set only, in order to choose the best model.⁶

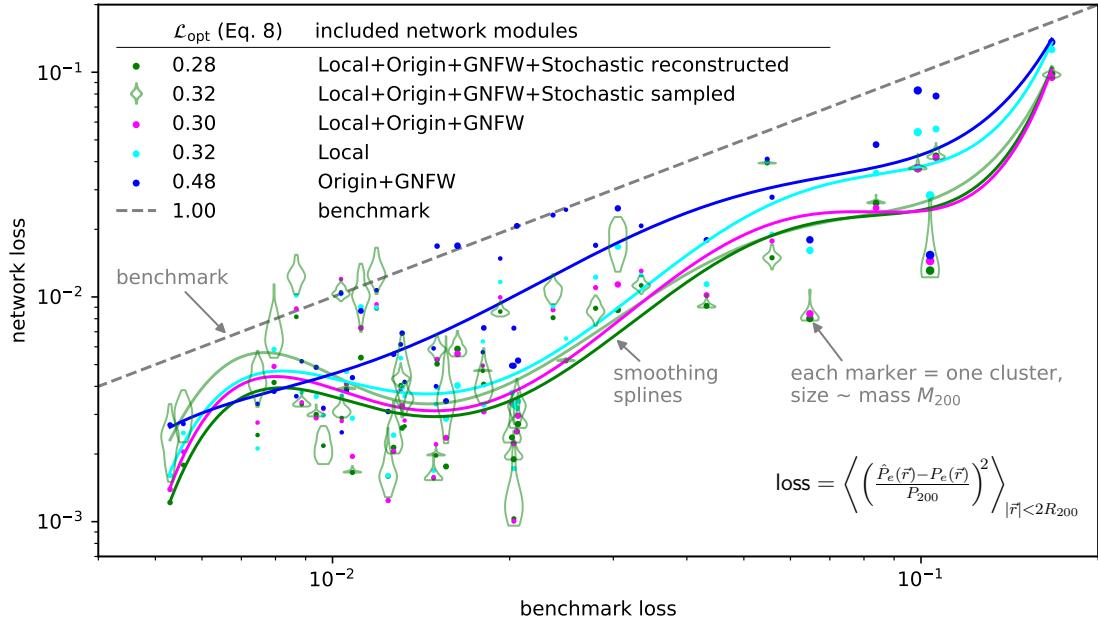


Figure 5.2: Network losses evaluated on testing set and compared against the GNFW benchmark model. Each data point is an individual cluster, the marker size indicating mass. For the version of the network including the *Stochastic* module, two modes of evaluation are possible: First, we can evaluate the reconstruction, i.e., the *Aggregator* receives the output μ of the *Stochastic* MLP (dark green points); second, we can randomly sample the VAE code, i.e., the *Aggregator* receives the vector \mathbf{a} drawn from a standard normal (light green violins). The lines are simple smoothing splines and only meant to guide the eye (the light green line corresponds to the means of the violins). The numbers in the legend’s first column are the performance metric \mathcal{L}_{opt} introduced in Eq. 5.8 (lower is better, benchmark $\equiv 1$).

5.4 Results and Discussion

In Fig. 5.2 we plot several network losses compared against the GNFW benchmark. Only correcting for mis-centering (blue) already gives a factor ~ 2 improvement over the use of cluster centers as identified by Rockstar. Likewise, only using the dark matter matter particles in the vicinity of the evaluation point⁷ (cyan) yields a further improvement. Combining the local information with the shifted GNFW profiles (magenta) performs better than *Local*-only by a few percent, the improvement being most pronounced in the high-loss regime. We conjecture that this could be

⁶Total compute cost is 13.4 (Tesla P100+9CPU) khr (1.09t CO₂e [269]) with a PyTorch [270] implementation.

⁷We find that $R \sim 300 \text{ kpc}/h$ is a good choice.

because the addition of the simpler GNFW model helps the network generalize in these relatively rare situations. Expectedly, the model including the *Stochastic* module (green) generally obtains lower reconstruction losses than the other models. The corresponding losses with random VAE samples (light green) are not much worse in most cases, although a larger training set would certainly help the network learn a more robust representation of the probabilistic component.

Naturally, we should ask whether our models are learning something trivial. We have checked that a more general spherically symmetric model, implemented as an MLP that takes as input $|\vec{r}|$ and the cluster scalars \mathbf{s}_α , does not perform more than a few percent better than the GNFW benchmark. Similarly, we find that a network using only the local density achieves more than twice the loss \mathcal{L}_{opt} compared to the *Local* network, demonstrating that the DeepSet is providing substantial information.

We have also experimented with adding further modules to the network. First, between *Origin* and GNFW we have inserted an MLP that uses the cluster \mathbf{s}_α , \mathbf{e}_α to account for deviations from spherical symmetry. We find no improvement from this modification. Second, we have constructed vector and scalar DeepSets operating on the cluster set $\{(\bar{q}_i^{(\alpha)}, \bar{v}_i^{(\alpha)})\}_{i \in \alpha}$ whose outputs were then passed to the *Aggregator*. Since these additional modules also do not yield any improvements, we conclude that the relatively large local regions contain enough information to infer the global properties of the cluster. It is important to appreciate that even these null results can tell us something physical, again a consequence of the interpretable, modular design.

In order to gain some further intuition, we now present two plots focusing on four example clusters from the testing set. These clusters were chosen at percentiles 25, 50, 75, and 100 in M_{200} to give a representative sample. We use the most powerful network for the predictions, corresponding to the green markers and violins in Fig. 5.2.

First, in Fig. 5.3 we show images of projected electron pressure. Note that the cutoff $r < 2R_{200}$ has been applied in all rows, so the projection depth varies with distance

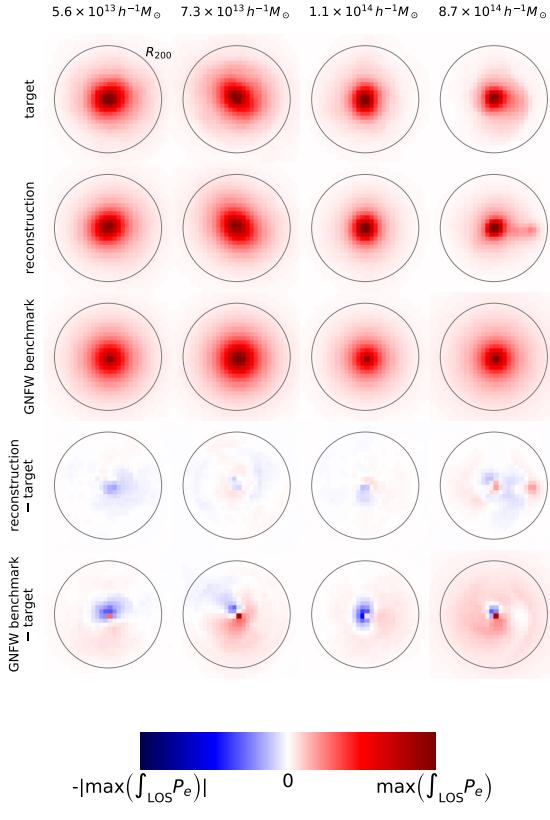


Figure 5.3: Projected electron pressure fields of four clusters. The projection directions were chosen randomly. Each ‘circle’ has radius $2R_{200}$. Cluster mass increases from left to right. The color scale is consistent within each column (i.e., for each cluster), normalized to the maximum pixel. The agreement between the target and the neural network prediction is good and always better than the GFW benchmark’s performance. Note the dipole pattern in the GFW residuals in the second and fourth column, indicating the effect of mis-centering. Deviations from spherical symmetry (compare the GFW prediction) are generally predicted in qualitative agreement with the target. However, the predicted field is generally too smooth, which is a common problem with neural networks. In Appendix D.1, we perform a qualitative comparison with results obtained using the CNN from [156].

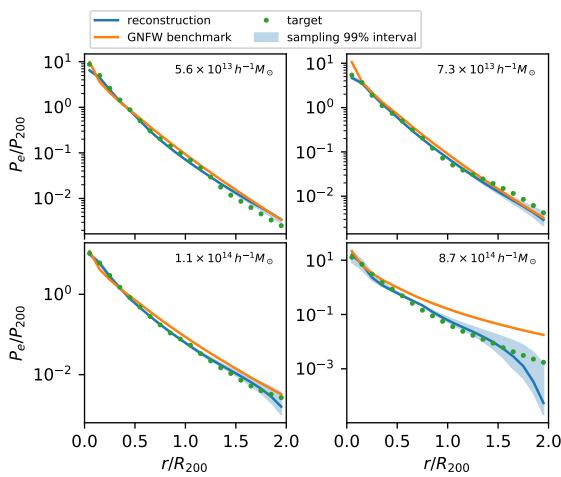


Figure 5.4: Electron pressure profiles of the same clusters as displayed in Fig. 5.3. The network predictions are generally in better agreement with the target than the GFW profiles. For all but the least massive objects the network appears to have learned a good estimate of its uncertainty (represented by the blue bands), with the predicted uncertainty increasing towards the cluster outskirts where substructure and the two-halo term start to become important. The neural network is able to predict a large dynamic range of electron pressure values, in particular the rare, dense cluster cores are predicted accurately.

from the cluster center. Visually, the spherically symmetric GNFW benchmark model provides a reasonable approximation to the target for all but the most massive object. The network predictions do generally pick up deviations from spherical symmetry in approximately the right direction. However, the network predicts fields that are noticeably smoother than the target; this is a commonly observed problem in similar tasks [246]. The most massive object (right column) appears to have undergone a recent merger event. Our network does realize this and provides a relatively good prediction, although the detailed structure of the electron pressure field around the secondary halo is not perfect. On the other hand, the GNFW benchmark struggles with this sample and overpredicts the pressure almost everywhere.

Second, in Fig. 5.4 we plot the electron pressure profiles corresponding to the same objects. This figure may be somewhat easier to interpret than Fig. 5.3, and it also gives us the opportunity to explore the probabilistic component of our network. We observe that generally the profiles predicted by our network are in better agreement with the target than the GNFW benchmark. It is noticeable that the GNFW profiles are generally too high around $r \sim R_{200}$, although we have carefully fitted them to the same data the network was trained on. This is a consequence of the fact that averaging in spherical shells and the locally evaluated loss function Eq. 5.7 do not commute in general. The stochastically sampled network predictions include the reconstructed profile in all four examples; this is a good indication that the network's representation of stochasticity is in fact a reasonable one. Furthermore, our expectation that the stochastic component is mainly responsible for a transfer of pressure between inner and outer parts of the cluster is confirmed. However, it should be noted that the predicted stochasticity in the profiles increases with cluster mass. This is difficult to reconcile with our argument that the stochasticity is primarily driven by the unresolved AGN activity, since AGN feedback is more efficient in altering the gas distribution in the shallower gravitational potential wells of lower-mass objects. This may be an

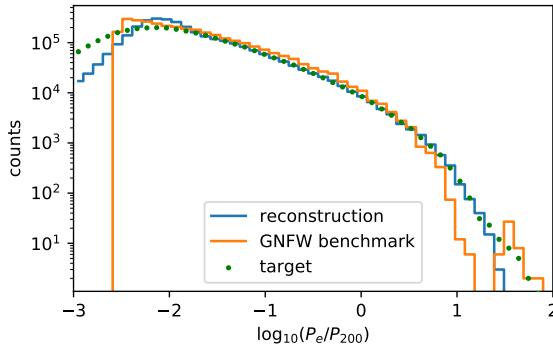


Figure 5.5: Histograms of electron pressure. We directly histogram the voxelized electron pressure values (c.f. Sec. 5.3). Thus, this figure is for illustrative purposes only, since the voxel sizes vary between different clusters (as they are fixed in units of R_{200}). Regardless, this plot is a useful diagnostic. The neural network is able to predict a large dynamic range of electron pressure values, performing much better than the GNFW benchmark. Low values are underpredicted because they contribute little to the loss function, while high values are overpredicted presumably due to their extreme sparsity.

indication that the primary driver of stochasticity is in fact the mass accretion rate, as was argued in [246]. Such an interpretation is further supported by the fact that the merger in the rightmost column of Fig. 5.3 seems to be the most challenging scenario for the network.

As a final result, we examine the histograms of electron pressure values, shown in Fig. 5.5. We observe that for the range $-1.8 \lesssim \log_{10}(P_e/P_{200}) \lesssim 1$ the network produces a histogram that is in almost perfect agreement with the target, while the GNFW benchmark shows some discrepancies. At low electron pressure, the network predictions cluster in a relatively narrow range $-2.3 \lesssim \log_{10}(P_e/P_{200}) \lesssim -1.8$ below which the histogram drops off compared to the target. This behavior is explained by the small contribution such small electron pressures give to the loss function. Similarly, for $\log_{10}(P_e/P_{200}) \gtrsim 1.2$, the network shows a lack of high-pressure values. This is most likely because predicting this small number of extreme peaks is a difficult objective given the small training set.

5.5 Future directions

We have developed a general method to construct interpretable models that predict continuous fields from a set of points while enforcing the underlying symmetries.

We have argued that the set-based approach introduced in this work has multiple advantages over CNNs that are commonly chosen for similar problems. Specifically, the direct operation on the simulation representation as a set of dark matter particles simplifies

- incorporation of inductive biases (such as the prominent role clusters play for the electron pressure field),
- modularization of the architecture into physically meaningful units,
- enforcement of underlying symmetries,
- fast training, through ‘sparsification’ of training data.

The application to cosmological structures demonstrates the power of our approach and opens up several directions of further investigation. The *Origin* module could possibly be almost directly incorporated in existing codes to improve the centering of spherically symmetric profiles. In order to get this work to a production-quality stage, training would need to be extended to lower-mass halos and non-zero redshifts, both of which should not present any major obstacles. Furthermore, it may be useful to combine the trained network with a CNN that handles the low-density regions outside the most massive objects. Such a combined architecture could then be used to predict cosmologically interesting summary statistics. Extensions to other baryonic fields beyond electron pressure should be straightforward and could potentially be easier since electron pressure is comparatively sparse. For fields that are better correlated with the dark matter density than is electron pressure, CNNs may benefit from the high locality. However, this would also apply to our set-based architecture, in which the cluster-scale modules could potentially be discarded in such cases. An interesting extension could involve transfer learning to sub-grid implementations different from IllustrisTNG.

In terms of interpretation, symbolic regression of e.g., the *Origin* module could provide useful insights and perhaps compact equations that would be more convenient

to implement in existing codes. A more detailed ablation study, e.g., on the importance of the various cluster-scale properties \mathbf{s}_α , could be interesting.

Beyond cosmology, we see potential use cases in irregular structures in condensed matter or in super-resolution atmosphere models from scattered meteorological measurements.

Acknowledgements. This work made use of the following software packages: `numpy` [271], `scipy` [272], `jupyter` [273], `sklearn` [274], `matplotlib` [275], `torch` [270], `torch_geometric` [276], and `astropy` [277].

Chapter 6

Cosmological constraints from HSC Y1 lensing convergence PDF

Abstract. We utilize the probability distribution function (PDF) of normalized convergence maps reconstructed from the Subaru Hyper Suprime-Cam (HSC) Y1 shear catalogue, in combination with the power spectrum, to measure the matter clustering amplitude $S_8 = \sigma_8 \sqrt{\Omega_m / 0.3}$. The large-scale structure's statistical properties are incompletely described by the traditional two-point statistics, motivating our investigation of the PDF — a complementary higher-order statistic. By defining the PDF over the standard deviation-normalized convergence map we are able to isolate the non-Gaussian information. We use tailored simulations to compress the data vector and construct a likelihood approximation. We mitigate the impact of survey and astrophysical systematics with cuts on smoothing scales, redshift bins, and data vectors. We find $S_8 = 0.860^{+0.066}_{-0.109}$ from the PDF alone and $S_8 = 0.798^{+0.029}_{-0.042}$ from the combination of PDF and power spectrum (68 % CL). The PDF improves the power spectrum-only constraint by about 10 %.¹

6.1 Introduction

Weak lensing surveys are rapidly catching up with the precision afforded by cosmic microwave background (CMB) experiments [e.g., 10, 8, 279]. Weak lensing directly probes the clustering of matter through statistical measurements of distortion of galaxy shapes. In recent years, weak lensing surveys have produced increasingly

¹based on Ref. [278], joint work with G. A. Marques, J. Liu, and M. Shirasaki.

precise measurements of the matter clustering amplitude S_8 [e.g., 280, 153, 281–283], showing hints of a discrepancy between galaxy lensing and CMB determinations, dubbed the S_8 -tension. Typically, these studies adopt two-point statistics — the two-point correlation function or the power spectrum.

While two-point statistics are sufficient to describe Gaussian fields such as the CMB, the large-scale structure that sources the weak lensing signal has undergone nonlinear growth and is thus far from Gaussian. Therefore, weak lensing non-Gaussian statistics have been proposed to extract complementary information.² Equally importantly, non-Gaussian statistics are affected by systematics differently than the two-point statistics. This is especially relevant for the S_8 -tension, since whether its origin stems from new physics or systematics is currently under hot debate. Non-Gaussian statistics will be particularly beneficial for upcoming surveys such as *Rubin* LSST [329], Euclid [330], and *Roman* [331]. These surveys will provide a high source density and hence probe deeper into the nonlinear regime.

In this work, we focus on the probability distribution function [PDF, 332–335] of the lensing convergence map³ from Subaru Hyper Suprime-Cam (HSC) Y1 data release [336, 337]. HSC is currently the deepest large-scale lensing survey and is thus considered a precursor of Stage-IV surveys. The PDF collects the amplitudes — but not shapes — of correlation functions of all orders. Thus, it is a highly non-Gaussian statistics and contains distinct information “far” from the power spectrum. Forecasts have shown promise in tightening the constraints on not only S_8 but also the neutrino mass sum and the dark energy equation of state [338–340, 156, 341, 342]. Formally, the PDF and moments are closely related. Moments can suffer from instabilities

²Popular non-Gaussian statistics include minima, peaks [284–292], voids [293], Minkowski functionals [294–300], Betti numbers [301, 302], persistent homology [303, 304], wavelets [305], scattering transform [306, 307], moments [308–310, 296, 297, 311, 312], higher-order correlation functions [313–318], density-split statistics [319], and convolutional neural networks [320–322]. A comparison of some of these in a forecast setting for Euclid has been carried out by [323], finding about similar performance for many of these statistics but limited to the Fisher approximation. An even more ambitious attempt are forward-model, field-level methods [e.g., 324–328]

³Also called mass map.

due to a few extreme pixels. This implies that, in practice, with a limited range of convergence values included in the PDF’s bins, the PDF can still be complementary to moments.

Our work is the first to obtain cosmological constraints from the lensing PDF with observations. While there exist analytic models of the PDF based on large-deviation theory [343–345] and the halo model [346, 347, 156], they are not yet at the level of accuracy and flexibility required to incorporate the complex survey configurations and systematics. Therefore, we use a large set of cosmological simulations tailored to HSC Y1 data to model the lensing PDF and its likelihood.

6.2 Methods

In this section, we provide a brief overview of the HSC Y1 data, the PDF and power spectrum measurements, the simulations suite, the likelihood, our blinding procedure, and systematics. More detail may be found in our companion paper on counts of lensing peaks and minima [348].

6.2.1 HSC Y1 Data

We use lensing convergence maps estimated from the HSC Y1 galaxy shapes catalogue [349]. After applying masks, the shear maps span 136.9 deg^2 in six spatially disjoint fields. We adopt galaxy redshifts determined using the MLZ code [350]. After applying redshift cuts of $0.3 < z_s < 1.5$ and restricting to sources with reliable shape measurements, we obtained a total number density of $\sim 17 \text{ arcmin}^{-2}$. We split source galaxies into four tomographic redshift bins with edges [0.3, 0.6, 0.9, 1.2, 1.5] and construct convergence maps via Kaiser-Squires inversion [351].

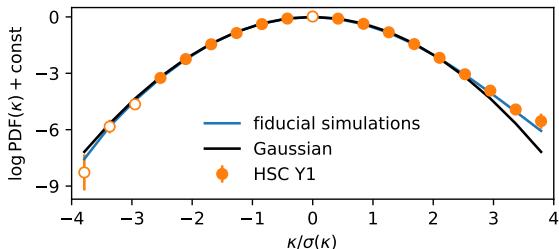


Figure 6.1: Example of the non-Gaussian lensing convergence PDFs from the HSC Y1 data (orange) and the fiducial simulation (blue), for the tomographic redshift bin $0.6 < z_s < 0.9$ and with a smoothing scale $\theta_s = 5$ arcmin. For most of the data points the error bars are invisible. Unfilled data points are not used in the analysis. The PDF's deviation from a Gaussian (black) provides the cosmological information to be exploited in this work.

6.2.2 PDF and power spectrum

The primary summary statistic we consider is the lensing PDF, calculated as histograms of pixels in normalized convergence maps. One crucial step in our computation is that we define the PDF over the signal-to-noise ratio, i.e., each convergence map is divided by its standard deviation before its pixels are histogrammed. This is important to remove information duplicated in the power spectrum. The PDF's non-Gaussian character becomes apparent in Fig. 6.1, where the tails deviate from a Gaussian (which is dominated by galaxy shape noise).

We measure PDFs on maps with 0.88 arcmin-sidelength pixels smoothed with Gaussian filters $\propto \exp(-\theta^2/\theta_s^2)$, where we choose $\theta_s = \{5, 7, 10\}$ arcmin to mitigate systematics while retaining non-Gaussian information. A Gaussian filter has optimal joint localization in configuration and reciprocal space and is thus a common choice [e.g., 292, 322, 323]. The presence of masking (where we set shear to zero) means that the filter may decrease the signal-to-noise ratio slightly, but no biases can arise since the mask is consistently implemented in data and simulations. For each smoothing scale, we histogram in 19 equally spaced bins between -4 and 4 (cf. Fig. 6.1). Of these, we remove the first three bins to minimize contamination by baryons according to tests described in Sec. 6.2.5 [also cf. 352]. Finally, we remove the 10th bin as otherwise the linear constraint from fixing the first three moments would render the PDF's auto-covariance almost singular.

We also consider the auto-power spectrum for each tomographic redshift bin, deconvolving the survey mask using the pseudo- C_ℓ method as implemented in `NaMaster` [353, 354]. We measure $C_\ell^{\kappa\kappa}$ in 14 logarithmically spaced bins in angular multipole $81 < \ell < 6580$. Of these, we remove $\ell < 285$ as Ref. [355] found unmodeled systematic errors for these scales in HSC Y1 data. Furthermore, we remove $\ell > 1000$ due to possible contamination by baryons. We note that $\ell_{\text{max}} \sim 1000$ is also comparable to the minimum smoothing scale of 5 arcmin used in the PDF data vector. This leaves us with four ℓ -bins. In contrast to previous two-point analyses [356, 357] we do not include cross-spectra between tomographic redshift bins. Upon unblinding the power spectrum-only posteriors we found that the highest tomographic redshift bin causes significant shifts in S_8 .⁴ Thus, we exclude the highest redshift bin in our analysis.

In summary, our raw data vector consists of the PDF with 135 dimensions (3 smoothing scales \times 3 redshifts \times 15 bins) and the power spectrum with 12 dimensions (3 redshifts \times 4 bins).

6.2.3 Simulations

We adopt two sets of N -body simulations in our analysis:

1. To model the covariance, we use 2268 realizations of the HSC Y1 footprint at a fiducial point ($S_8 = 0.791$, $\Omega_m = 0.279$), generated from the 108 quasi-independent [358] full-sky maps of Ref. [359].
2. To model our statistics, we use simulations at 100 different cosmological models with varying values of S_8 and Ω_m , as introduced in Ref. [360]. For each cosmology, 50 quasi-independent realizations are generated by randomly placing observers within the periodic box. These simulations are used to construct emulators for the

⁴This may be due to deficiencies in our simulations; for example, the simulations use lower resolution at high redshifts which may be not sufficient. Alternatively, the highest source redshift bin may be subject to photo- z calibration errors [282, 283].

mean theory prediction; thus, they are designed to have better accuracy than the fiducial simulations.

For each of these 7768 ($=2268+100\times50$) realizations, mock galaxy shapes are generated following Refs. [361–363]. Convergence map generation and summary statistic measurements are performed on these simulated mocks identically to as done for the real data.

6.2.4 Likelihood and inference

For the PDF, to reduce the size of the data vector and Gaussianize its likelihood, we score-compress the logarithmic PDF under the approximation of a Gaussian likelihood [MOPED, 364].⁵ This reduces the number of PDF bins from 135 to only 2, corresponding to the number of parameters (S_8 and Ω_m). We then construct at each of the 100 cosmological models a 2×2 covariance matrix, using the 50 realizations. Finally, we build emulators of both the compressed PDF and the cosmology-dependent inverse covariance using Gaussian processes.⁶ If we use a cosmology-independent covariance the S_8 -posterior tightens and its peak is almost unchanged. However, with that choice the ranks plot discussed below would indicate an overconfident posterior.

The power spectrum data vector is small enough and its distribution is known to be close to Gaussian. Therefore, we apply no data compression to the power spectrum. We build a Gaussian process emulator to model the power spectrum and estimate a cosmology-independent covariance matrix using the fiducial simulations. To jointly analyze the compressed PDF and the power spectrum, we approximate the cross-correlation between them as constant and estimate it from the fiducial simulations.

⁵We note that upon combining the PDF with the power spectrum, the compression is likely sub-optimal; future work could investigate more rigorous ways to preserve a maximum of information. Of course, the choice of likelihood to compute the score cannot induce biases, only information loss.

⁶Analytic methods to compute the covariance have been presented in Refs. [156, 365].

We adopt uniform priors on our parameters S_8 and Ω_m , in intervals $[0.5, 1.0]$ and $[0.2, 0.4]$, respectively. Our prior is well-covered by the available simulations. Markov chain Monte Carlo (MCMC) sampling is performed using `emcee` [366, 367].

We validate our likelihoods with a “ranks plot” [e.g., 368] shown in Fig. 6.2. The rank statistic is a simple calibration test which relies on the fact that for a valid likelihood random draws from the posterior should be statistically indistinguishable from the true parameter vector. To construct the ranks plot, we run MCMC on realizations drawn from the 30 cosmology-varied simulations within our prior. To cleanly separate our training and test sets, all emulators are rebuilt without the test cosmology. We then order each Markov chain by its S_8 values and find the rank of the true S_8 . For a valid likelihood, it should be impossible to statistically distinguish the true S_8 from randomly selected Monte Carlo samples, so the ranks should follow a uniform distribution. If the posterior is over-confident (under-confident), the histogram exhibits a U-shape (inverse U). However, because we only have a relatively small number of simulated cosmologies available within our prior, a perfectly uniform distribution may not be possible to attain. Indeed, Fig. 6.2 shows an approximately uniform distribution for all data vector choices (PDF, power spectrum, and them jointly), except for spikes at the edges for the power spectrum and joint statistic. These spikes are attributable to a small number of simulations near the prior boundary where the power spectrum emulation appears to work less well. We do not expect these to affect our results as they have little overlap with our final posterior and the problem is mostly due to underestimated tails which are less important for the confidence levels typically considered.

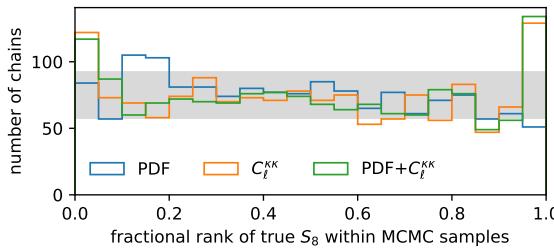


Figure 6.2: Validations of our likelihoods. The histograms are over hundreds of Markov chains, run with realizations from the cosmo-varied simulation set as “observation”. The x -axis is the position of the simulation’s true S_8 value in the ordered chain. A uniform distribution is expected for a valid likelihood. The grey band represents the 2σ interval approximating the bin counts as Poissonian.

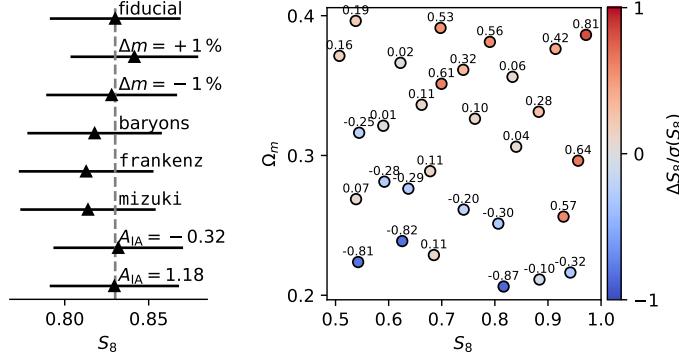


Figure 6.3: *Left:* Estimate of systematic errors. We contaminate the $\text{PDF} + C_\ell^{\kappa\kappa}$ data vector with various systematics as described in the text. No systematic exceeds $0.45\sigma(S_8)$. *Right:* Assessment of training set and emulator quality. Each point is one of the cosmologies for which cosmo-varied realizations exist. The bias indicated is estimated by averaging over chains run on 25 realizations, while the emulator is trained with the other 25 realizations and without including the cosmology where the inference is performed.

6.2.5 Systematics

We study potential biases caused by effects not included or wrongly described in the simulated model by running inference on fiducial mocks contaminated with the following methods:

- We simulate miscalibration of multiplicative bias by shifting it by $\pm 1\%$, corresponding to the uncertainty on the mean quoted in Ref. [369].
- We assess the impact of photometric redshift uncertainty by generating fiducial mocks with source redshifts from two other codes, `frankenz` and `mizuki`.

- We model the impact of baryonic effects using the κ TNG simulations [370] — two sets of convergence maps with and without baryons. We contaminate the fiducial mocks with the ratio of hydrodynamic to dark matter-only data vectors.
- We estimate the impact of intrinsic alignments similarly to the baryons case by multiplying the fiducial data vectors with a fixed ratio which we obtain from mocks with nonlinear intrinsic alignments [371]. We choose alignment amplitudes, $A_{\text{IA}} = \{-0.32, 1.18\}$, comparable to the 1σ constraints in Refs. [356, 357].

We show the resulting biases on S_8 for these systematics in the left panel of Fig. 6.3 (also see the more comprehensive discussion in Ref. [348]). The most constraining joint PDF+ $C_\ell^{\kappa\kappa}$ data vector is used. We do not find biases exceeding 0.45σ in S_8 .

In addition, we also consider potential biases caused by imperfections in the simulations and emulators:

- The mean recovered S_8 on the fiducial simulations is about 1σ higher than the input, indicating a systematic difference with the cosmology-varied simulations. This is likely due to the lower fidelity of the fiducial simulations. With this in mind, the observed 1σ bias can be considered an upper bound on biases caused by resolution effects in the N -body simulations.
- To assess biases caused by the limited training set size, at each cosmology we divide the realizations into 25 samples each for training and testing. We match the realization indices (i.e., observer positions) between cosmologies so as to maximize statistical independence between training and testing samples. In addition, all realizations at the test cosmology are removed from the training. The right panel of Fig. 6.3 shows the average biases on S_8 within our prior range, in units of the standard deviation. The bias is computed using the mode of the marginalized S_8 posterior, to minimize the effect of the prior which can shift the mean. The bias is a few ten percent for most of the points, except at the edges where the emulator

quality decreases. Since the described test aggressively reduces the training set size, our bias estimates are likely a conservative upper bound.

6.2.6 Blinding

To help build best practice for non-Gaussian statistical analyses with Stage-IV surveys [e.g., 323], we follow a three-step blinding procedure in our analysis:

1. We build the mock generation pipeline using one random realization from the fiducial model as “observation”. This includes shear bias correction, convergence map reconstruction, masking, in-painting, and summary statistics measurements.
2. We construct the inference pipeline using the ranks plot and similar tests. This includes data compression, covariances, emulators, and MCMC sampling.
3. We select smoothing scales, redshift bins, and cuts on data vectors to minimize the impact of systematics.
4. First unblinding: B-modes. We compare the power spectra and PDFs measured in B-mode maps of the HSC Y1 data and our fiducial simulation.⁷ We validate that the two distributions are statistically consistent.
5. Second unblinding: power spectrum. We apply inference to the measured HSC Y1 power spectrum and compare our results internally between different redshift bins and to the official HSC Y1 analyses by Refs. [356, 357]. We discover systematics in the highest tomographic redshift bin and remove it from our final analysis.
6. Third unblinding: PDF. We unblind the PDF-only posterior during an invited presentation at the HSC weak lensing working group telecon. The initial posterior was approximately uniform. This failure was due to a bug in computing the

⁷B-mode maps are built by rotating galaxy shapes by 45 degrees.

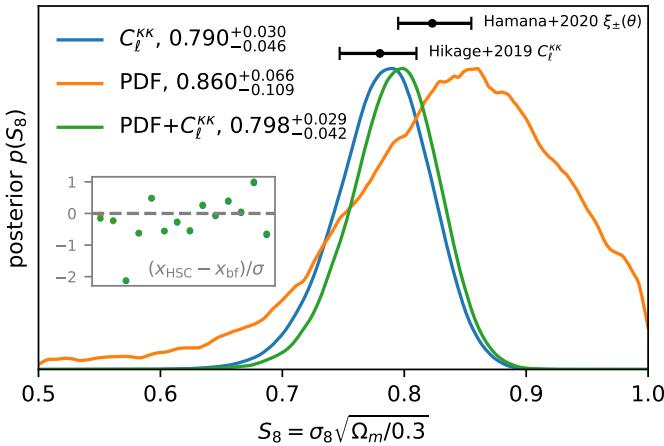


Figure 6.4: S_8 posteriors from different data vector choices. Since the posteriors are slightly asymmetric, we quote their modes as well as the symmetric 68% confidence interval. For comparison, we include 68% intervals from Refs. [356, 357]. Note that these analyses differ in various details. The inset compares the best-fit model to the data (PTE = 0.86). We remind the reader that the PDF does not contain information on the variance.

standard deviations of the HSC data maps.⁸ The bug did not affect our previous steps, so once we resolved it we did not change any other parts of the pipeline.

7. After the initial submission of this paper, we learned that there are two additional multiplicative bias parameters arising from galaxy size selection and redshift-dependent responsivity corrections. Post-unblinding, we updated our analysis to account for this total multiplicative bias. The correction is very small and hence did not change our conclusions. However, it made our power spectrum-only results more consistent with the official analysis [356].

Since the HSC Y1 data is already public our blinding procedure is rather an honor system.⁹

6.3 Results

We show the S_8 posteriors obtained from the power spectrum, the PDF, and the two combined in Fig. 6.4. In all cases, the constraint on Ω_m is prior-dominated and thus

⁸Thanks to the feedback from the HSC weak lensing working group, we learned a valuable lesson of comparing the PDFs before moving on to inferences. They also pointed out that a more sophisticated blinding policy, as typically employed by collaborations, would have flagged this issue before unblinding.

⁹During our study, we also became aware of other recently completed non-Gaussian statistical analyses using the same dataset [322, 372]. To avoid confirmation bias, we refrained from reading these papers until we finalized our results.

not shown here for brevity. For reference, we also show the 68 % intervals from the official HSC Y1 two-point analyses with the power spectrum [356] and the two-point correlation function [357].

Our two-point result (blue) agrees well with the previous analyses in terms of mean value. However, our posterior is wider: $\sigma \simeq 0.042$ compared to $\sigma \simeq 0.03$ in the previous two-point analyses. This is expected as these analyses differ in many respects, such as additional cross-spectra between tomographic redshift bins, choice of priors, treatment of systematic errors, data vector binning, and scale cuts.

Our PDF-only posterior (orange) shows a clear detection of S_8 , or specifically *nonlinear* clustering, at 9σ . This is after removing variance information from the PDF and thus almost independent of the two-point results. The mean S_8 value is slightly higher than the one from $C_\ell^{\kappa\kappa}$ but statistically consistent. Upon combining the PDF and $C_\ell^{\kappa\kappa}$, we observe a slight upwards shift in the mean S_8 and a $\approx 10\%$ tightening of the posterior. This shows that the PDF indeed contains complementary information to the power spectrum.

In many aspects, our results match the CNN analysis from Ref. [322], including the S_8 posterior and an $\mathcal{O}(10\%)$ improvement compared to the power spectrum.

6.4 Conclusions

In this work, we obtained the first cosmological constraints using the probability distribution function of weak lensing convergence. We built convergence maps using the HSC Y1 shear catalogue, constructed a likelihood based on tailored numerical simulations, and validated that known survey and astrophysical systematics are under control after applying cuts on smoothing scales, redshift, and data vectors. We designed and followed a three-step blinding procedure to minimize confirmation bias. We make our inference code publicly available at this URL.

The PDF improves the power spectrum-only constraint by about 10%. For the clustering amplitude, we obtained $S_8 = 0.860_{-0.109}^{+0.066}$ and $S_8 = 0.798_{-0.042}^{+0.029}$ from the lensing PDF alone and the combination of PDF and power spectrum, respectively (68% CL). We computed the PDF on convergence maps normalized by their standard deviation, hence maximally removed two-point information from the PDF. Our results are consistent with previous analyses on the same data and show that the PDF provides additional information not contained in the power spectrum.

We find no tension between the S_8 inferred from HSC Y1 lensing and from primordial CMB measurements.

Future work could investigate alternatives to the probably lossy compression step as well as the forward modeling of systematics to allow inclusion of smaller scales. Additionally, Stage-IV data will be sensitive to cosmological parameters beyond S_8 , for example, the neutrino mass sum. The PDF could be instrumental in complementing two-point statistics in order to optimally constrain such model extensions.

Acknowledgements. We thank Sihao Cheng, Will Coulton, Daniela Grandón, Kevin Huffenberger, Xiangchong Li, Surhud More, Ken Osato, David Spergel, Sunao Sugiyama, and Masahiro Takada for useful discussions. We thank Joachim Harnois-Déraps for sharing the IA mocks. We thank the organizers of the Kyoto CMBxLSS workshop, during which this work was completed. The work of LT is supported by the NSF grant AST 2108078. This work was supported by JSPS KAKENHI Grants 23K13095 and 23H00107 (to JL), 19K14767 and 20H05861 (to MS). This work was supported in part by the UTokyo-Princeton Strategic Partnership Teaching and Research Collaboration. This manuscript has been authored by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics. The authors are pleased to acknowledge that the work reported on in this paper was substantially performed using the Princeton Research Computing resources at Princeton University which is consortium of groups led by the Princeton Institute for Computational Science and Engineering (PICSciE) and Office of Information Technology's Research Computing.

Chapter 7

Neutrino mass constraint from an Implicit Likelihood Analysis of BOSS voids

Abstract. Cosmic voids identified in the spatial distribution of galaxies provide complementary information to two-point statistics. In particular, constraints on the neutrino mass sum, $\sum m_\nu$, promise to benefit from the inclusion of void statistics. We perform inference on the CMASS NGC sample of SDSS-III/BOSS with the aim of constraining $\sum m_\nu$. We utilize the void size function, the void-galaxy cross power spectrum, and the galaxy auto power spectrum. To extract constraints from these summary statistics we use a simulation-based approach, specifically implicit likelihood inference. We populate approximate gravity-only, particle neutrino cosmological simulations with an expressive halo occupation distribution model. With a conservative scale cut of $k_{\max} = 0.15 \, h\text{Mpc}^{-1}$ and a *Planck*-inspired ΛCDM prior, we find upper bounds on $\sum m_\nu$ of 0.43 and 0.35 eV from the galaxy auto power spectrum and the full data vector, respectively (95 % credible interval). We observe hints that the void statistics may be most effective at constraining $\sum m_\nu$ from below. We also substantiate the usual assumption that the void size function is Poisson distributed.¹

¹based on Ref. [373], joint work with E. Massara, A. Pisani, C. Hahn, D. N. Spergel, S. Ho, and B. Wandelt.

7.1 Introduction

The Universe’s ability to provide glimpses into experimentally inaccessible conditions has a long history, including the deduction of the laws of Gravity and the discovery of helium. In the present day, cosmology offers a unique view on the properties of neutrinos which are amongst the last unknowns in the standard model of particle physics. First evidence for a non-zero neutrino mass sum, $\sum m_\nu$, came from the solar neutrino problem [374–376]. Subsequently, oscillation experiments provided proof that neutrinos must have mass [377–381] and established the lower bounds of 0.06 and 0.1 eV in the normal and inverted hierarchy, respectively. The terrestrial experiment KATRIN currently sets an upper bound of ≥ 0.8 eV [382].² However, the strongest upper bounds are already provided by cosmological data, the primary CMB alone giving 0.38 eV [8], for example. It will be one of cosmology’s primary goals in the coming decade to tighten this bound and eventually detect neutrino mass.

One of the natural regimes to look at to constrain $\sum m_\nu$ are extremely underdense regions, cosmic voids [383–386]. As the cold dark matter (CDM) flows out of the voids and into filaments and clusters, neutrinos are more smoothly distributed. Thus, the neutrino/CDM ratio is higher in the voids and lower in the clusters. These qualitative considerations have spawned considerable theoretical interest in the use of void properties to constrain $\sum m_\nu$. This includes simulated data vector-level investigations [387–392] as well as forecasts [393–395]. The forecasts find promising error bars on $\sum m_\nu$, albeit under simplifying assumptions. Voids may be the first regime in which non-linear signatures of massive neutrinos will be observed [also c.f. 396].

Being large objects, voids had to wait for the era of relatively deep, large-volume surveys with approximately uniform selection function to be statistically usable. While

²The KATRIN bound is on $m_\beta^2 \equiv \sum_\nu |U_{e\nu}^{\text{PMNS}}|^2 m_\nu^2$, so it only equals a bound on $\sum m_\nu$ for a special, experimentally excluded, choice of the PMNS matrix and the mass hierarchy. In general, the bound is weaker.

the original detections focused on individual objects [397–401], which already contain cosmological information [402], we are now able to utilize catalogs of hundreds and thousands of voids [403–408] to perform precision cosmology with void shapes [409–417] and sizes [418, 419].

In this work, we use voids identified in the CMASS sample of the Sloan Digital Sky Survey (SDSS)-III Baryon Oscillation Spectroscopic Survey (BOSS) [420–422] to place constraints on $\sum m_\nu$. The void statistics we consider are the void size function (VSF) and the void-galaxy cross power spectrum. We combine these with the usual galaxy auto power spectrum multipoles which by themselves already place a tight upper bound on $\sum m_\nu$ (through the suppression of matter power below the neutrino free-streaming scale) [e.g., 423, 424].

Since voids can be considered anti-halos [425], a popular model for the VSF descends from Press-Schechter theory and the excursion set formalism [426–428], with slight modifications [429–439]. While the void-galaxy correlation function can be used for cosmological purposes without explicit knowledge of the void profile through the Alcock-Paczynski test and redshift space distortions [440–442], the modeling of the profile itself has also been considered [443–447, 395].

However, all analytic approaches to modeling void statistics are problematic for our purposes. First, it is difficult to construct a consistent galaxy bias model across the different statistics comprising the data vector. Second, the calibration of analytic models typically did not utilize large simulations with varied neutrino mass. Third, existing models typically apply only to an aggressively cleaned subset of the entire void catalog, potentially leading to appreciable losses in constraining power.

Therefore, we choose to work in a simulation-based framework. Our simulations are based on particle-neutrino, approximate gravity-only `FastPM` [448, 449] realizations, in which we place galaxies through an expressive halo occupation distribution model

(HOD) [450, 130, 451]. We then post-process the galaxy catalogs to generate light cones incorporating survey realism.

The likelihood analysis with these simulations is a non-trivial problem. A popular approach is to build emulators of the mean data vector and perform the analysis under the assumption of a usually Gaussian likelihood where the covariance matrix is estimated from simulations. However, this approach turns out to be challenging for our problem. First, constructing an emulator in a 17-dimensional space (6 cosmology, 11 HOD) is quite difficult, especially given a feasible number of simulations. Second, the assumption of a Gaussian likelihood is wrong. We demonstrate in Sec. 7.4.5 that the VSF is very close to Poisson distributed (as long as bins are chosen wide enough, as would be naively expected), but modeling its covariance with the void-galaxy cross power spectrum and the galaxy auto power spectrum is difficult.

For these reasons, we opt for an implicit-likelihood³ approach [452]. This formalism uses neural networks to approximate functions that can be converted into posteriors. In general structure, this work is therefore similar to the **SIMBIG** papers [453, 454], but it differs in almost all details (statistics, simulations, objective, HOD, code). The resulting complementarity will therefore be useful to assess the state of implicit likelihood inference in galaxy clustering cosmology.

The rest of this paper is structured as follows. Sec. 7.2 describes our simulation pipeline. Sec. 7.3 contains details on the data vector and the inference procedure. Sec. 7.4 collects our results and their interpretation. We conclude in Sec. 7.5. The appendices contain additional material as well as information about data and code availability.

³likelihood-free, simulation-based.

7.2 Simulations

7.2.1 Cosmological prior

Since our objective is $\sum m_\nu$, we place a tight prior on ΛCDM . For this, we use the posterior from the *Planck* [8] primary CMB analysis.⁴ Specifically, we use the chains run with fixed $\sum m_\nu$ and measure the mean and covariance matrix in the five “CMB parameters” ω_c , ω_b , $\log A_s$, n_s , and θ_{MC} . In these parameters the posterior is close to Gaussian and we approximate it as such. To ensure the robustness of our conclusions, we inflate the *Planck* error bars on cosmological parameters by a factor of two. For $\sum m_\nu$, we choose a flat prior between 0 and 0.6 eV, the upper boundary being motivated by preliminary tests in which we established a sensitivity of the order $\sigma \sim 0.2$ eV. We assume three neutrino species with degenerate masses.

Of course, the primary CMB’s information leads to some correlation between $\sum m_\nu$ and the CMB parameters. This correlation is not included in our prior. However, given the sensitivity of the data used (compared to *Planck*), these residual correlations have a relatively small effect. For example, projecting the $\sum m_\nu - \omega_c$ correlation in the *Planck* posterior to our upper prior boundary of 0.6 eV, we obtain a shift $\Delta\omega_c/\sigma_{\text{EFT}} \sim 0.25$ where σ_{EFT} is the error bar obtained from the EFTofLSS analysis of BOSS [455] (of which we only use a subset). From these considerations, it also follows that our results do not depend strongly on the precise choice of ΛCDM prior.

We draw from the cosmological prior using an open quasi-random sequence. In contrast to popular sampling methods such as latin hypercube or Sobol, an open sequence does not require initial knowledge of the total number of samples. Our sequence is constructed by taking integer multiples of a vector whose elements are powers of a generalized golden ratio.⁵ In hindsight it turns out that our results are insensitive to taking out random subsets from the sequence (c.f. appendix F.6),

⁴plikHM-TTTEEE-lowl-lowE

⁵<http://extremelearning.com.au/unreasonable-effectiveness-of-quasirandom-sequences>

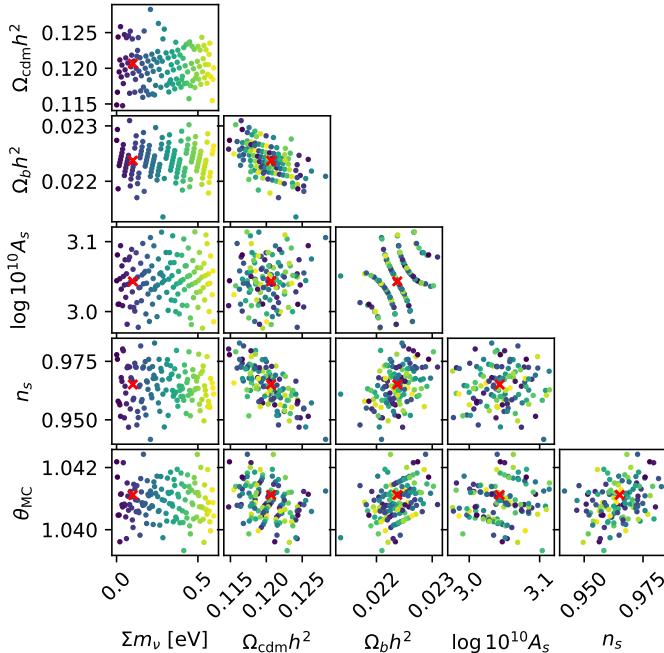


Figure 7.1: The cosmological parameter values sampled. We adopt a prior on ΛCDM that has the same shape and twice the size (in σ) as the *Planck* posterior. We sample this prior with an open quasi-random sequence designed to yield low-variance estimates of integrals. The color scale correlates with $\sum m_\nu$. The red marker indicates the position of our fiducial simulations.

indicating that a simple pseudo-random sampling would have been sufficient. This is likely due to the aforementioned compactness of the prior compared to the data's sensitivity.

7.2.2 Cosmological simulations

We run 127 simulations with varied cosmologies and 69 at a fiducial cosmology, illustrated in Fig. 7.1.⁶ We choose the fiducial cosmology close to the mean of the ΛCDM prior, with $\sum m_\nu = 0.1 \text{ eV}$. The cosmo-varied simulations share the random seed. This is because the decision to adopt an implicit-likelihood approach was only made after we encountered severe challenges with the conventional approach, as described in the introduction. However, as we shall see below, the simulations are large enough that sufficient quasi-independent data vectors can be generated.

⁶We attempted 130/71, but some runs failed. As discussed before, the fact that due to the failures we do not sample the quasi-random sequence strictly sequentially does not affect our results.

After generating a cosmological parameter vector in the CMB parameters we replace θ_{MC} by the Hubble constant using **CAMB** [456, 457]. We then produce power spectra at $z = 99$ using **CLASS** [458] and **REPS** [459, 460].

We run particle neutrino gravity-only simulations using the approximate **FastPM** solver. We choose a box size of $2.5 h^{-1}\text{Gpc}$ with 2800^3 CDM particles, leading to a minimum resolved halo mass of $\sim 1.3 \times 10^{12} h^{-1}M_\odot$ that is sufficient for the CMASS LRG sample. With regard to the neutrino options in **FastPM** we follow Ref. [449].⁷ In particular, we follow their recommendation to increase the number of early time steps for larger neutrino masses. Specifically, at $\sum m_\nu = 0$ we take seven logarithmic steps (in scale factor) between $z = 99$ and $z = 19$, followed by twelve linear steps until $z = 0.68$. Afterwards, we take twenty steps until $z = 0.44$ during each of which we write a snapshot of CDM particles to disk. At higher neutrino masses we insert up to twelve additional logarithmic steps before $z = 79$. The seemingly large number of twenty snapshots was established in our preliminary tests during which we saw slight differences between ten and twenty snapshots (in the summary statistics considered and for a single **FastPM** run) and thus decided to err on the side of caution. Note that a recent study supports the accuracy of **FastPM** for a task similar to ours, albeit with some important differences [461].

Each simulation takes about 90 minutes on 70 40-core nodes of the Tiger machine at Princeton.

7.2.3 Galaxies

We fine-tuned the method to populate CDM snapshots with galaxies through preliminary tests. Specifically, we performed global optimization over a large and partly discrete space of halo occupation distributions considering two objectives: (1) the power spectrum multipoles of galaxies placed in Quijote simulations [462]; (2) the

⁷`every_ncdm=4, n_side=4, n_shell=10`

VSF of the CMASS data. The first step was intended to identify degrees of freedom that are necessary to correct for potential approximation errors in `FastPM`, while the second test was primarily meant to test our simulations' fidelity. The optimization problems were solved with `optuna` [144]. In the following, we briefly describe the halo occupation distribution model (HOD) resulting from these preliminary tests. More detail can be found in appendix F.1.

We identify halos in the CDM snapshots using `Rockstar` [264, 265], which in our preliminary tests performed better than the friends-of-friends finder shipped with `FastPM`.

Then, galaxies are assigned stochastically to halos using an HOD. Besides the usual five-dimensional model parameterized by M_{\min} , $\sigma_{\log M}$, M_0 , M_1 , α [463], we introduce six additional degrees of freedom.

Although galaxy assembly bias [464–466] has been argued to be not necessary to describe the clustering of CMASS galaxies [467–469], we decide to be conservative by adding galaxy assembly bias parameterized by P_1 and a_{bias} . Furthermore, we add velocity bias [470–473] parameterized by η_{cen} and η_{sat} . Finally, we introduce redshift dependence to M_{\min} and M_1 , parameterized by $\mu(M_{\min})$ and $\mu(M_1)$. One advantage of having these slopes as free parameters is that we know them to be relatively close to zero, enabling useful sanity checks on any posteriors.

The resulting 11-dimensional HOD parameterization is quite similar to, e.g., Refs. [474, 453, 454].

We populate the cosmo-varied simulations with galaxies according to HOD parameters drawn from the priors given in appendix F.1. For the simulations at the fiducial cosmology we only populate with a single HOD. We choose the HOD parameters used for the fiducial mocks based on preliminary inference runs using the VSF only. It turns out that these parameters are not very close to best-fit when considering the entire data vector; generating new mocks closer to the best-fit point may increase the

efficiency of the compression step described below. However, the difference in HOD parameters cannot cause biases since the fiducial mocks are not used in constructing the likelihood.

7.2.4 Light cones

We use the cuboid remapping code `BoxRemap`⁸ [475] to deform our simulated cubes to the CMASS NGC geometry. It turns out that there are two possible choices of remapping and we use both (as part of the augmentation scheme discussed below).⁹

When projecting galaxies onto the light cone, we extrapolate their positions from the snapshots using the host halo velocities (using the stochastic galaxy velocities would weaken the correlation between centrals and satellites). The resulting corrections are small thanks to the large number of available snapshots.

After mapping galaxies to the light cone, we apply all angular masks and approximately mimic fiber collisions using the procedure described in Ref. [453].

In contrast to some other works, we downsample the galaxy field predicted by the HOD to the data’s density $n(z)$ (only, of course, if the simulation contains more galaxies at the given redshift). This downsampling is performed iteratively in conjunction with the implementation of fiber collisions so that both are self-consistent. Our implementation performs any necessary downsampling regardless of host halo properties; future work could take a prediction for stellar mass into account.

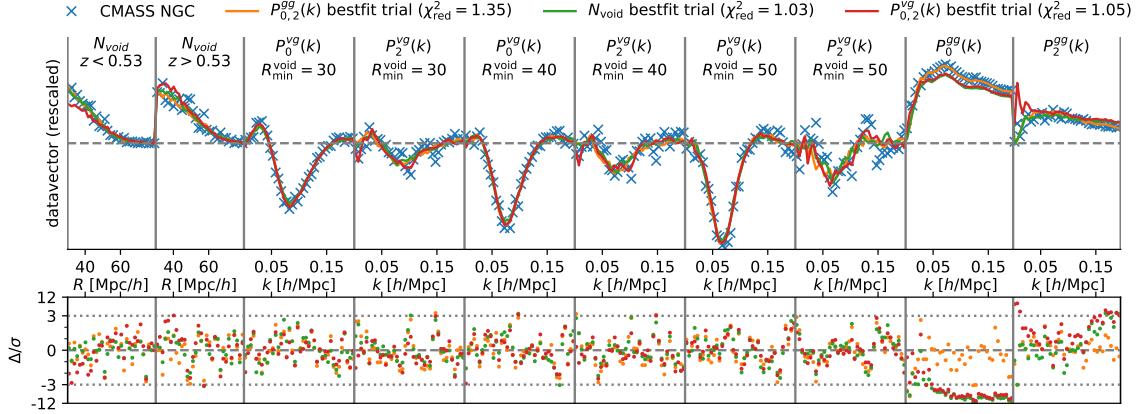


Figure 7.2: Illustration of the data vector considered. In the *top panel*, the vertical axis has been scaled such that the different data vector components are well visible and the power spectra are plotted as $kP(k)$. The solid lines display best-fit mocks, considering separate parts of the data vector (the notation $P_{0,2}^{\text{gg}}$ denotes the combination of monopole and quadrupole of the galaxy auto power spectrum, etc.). The χ^2 was computed under the approximation of a Gaussian likelihood and the mocks were averaged over eight augmentations (no interpolation/emulation was performed). The fact that, e.g., the model with best-fit VSF still reproduces the other parts of the data vector reasonably well, as well as the reduced χ^2 values close to unity, indicate our mocks’ high fidelity. Note that we do not use $k < 0.01$ and $k > 0.15$ in our baseline analysis and these scales were not included in identifying the best-fit models shown here. The *bottom panel* shows the residuals in units of estimated standard deviations. For better visibility, the y -axis has a linear-logarithmic scale with transition indicated by the dotted lines.

7.3 Inference

7.3.1 Data vectors

We use the northern galactic cap (NGC) part of the CMASS sample. The southern part (SGC) is smaller and we do not expect dramatic improvements from its inclusion. Since our focus is on better understanding the impact of void statistics, rather than the tightest possible bounds on neutrino mass, we ignore the SGC for simplicity. Similarly,

⁸<http://mwhite.berkeley.edu/BoxRemap>

⁹The two cuboids have sidelengths $\{1.412, 1, 0.7071\}$ and $\{1.7321, 0.8165, 0.7071\}$ (in units such that $2.5 h^{-1}\text{Gpc} \equiv 1$). They are generated by the matrices $\{\{1, 1, 0\}, \{0, 0, 1\}, \{1, 0, 0\}\}$ and $\{\{1, 1, 1\}, \{1, 0, 0\}, \{0, 1, 0\}\}$, respectively.

we do not include the LOWZ sample; its lower volume makes it less suitable for void science.

We cut galaxies into the redshift interval $0.42 < z < 0.7$ and map them to comoving space using a fixed $\Omega_m = 0.3439$. Voids are identified using the VIDE code [476] which is based on ZOBOV [477] and works by Voronoi tessellating the galaxies and then applying a watershed algorithm to find contiguous density minima. We use the “untrimmed” catalog computed by VIDE as it does not require arbitrary assumptions.

While many different void finders exist [e.g., 478–480], prior work suggests that shape-agnostic void finders such as VIDE yield voids with better constraining power on $\sum m_\nu$ than spherical finders [395]. Future work could investigate the influence of void definition on signal-to-noise.

Galaxy auto power spectra $P_\ell^{gg}(k)$ and void-galaxy cross power spectra $P_\ell^{vg}(k)$ are computed using `nbodykit` [481, 482] and `pypower`,¹⁰ reducing variance with FKP weights [483] and correcting for observational systematics using the provided weights (except for fiber collisions, of course) [484]. We only utilize the systematics weights when computing P^{gg} . In the case of P^{vg} , we find no significant change in posteriors when using the galaxy weights, consistent with Ref. [412]. In the case of void identification, there is no guarantee that the obvious method to incorporate the systematics weights would yield cleaner voids. Given the relatively large void sizes considered in this work, we do not expect significant contamination by unmodeled survey systematics, but suggest that this point may warrant future work. Galaxy randoms are taken from the public catalogs. Void randoms are constructed by taking a large catalog of voids from many different mocks to choose angular positions and constructing a kernel density estimator in redshift matched to the specific void catalog. This procedure ensures that the randoms are consistent with a given cut in void radius

¹⁰<https://pypower.readthedocs.io>

since voids of different sizes have somewhat different angular distributions due to the survey mask.

We consider the VSF in 32 linearly spaced effective radius bins between 30 and $80 h^{-1}\text{Mpc}$. The minimum radius cut is well above the mean tracer separation and thus we expect contamination by Poisson voids to be small [485, 486]. We split voids for the VSF into two redshift bins, separated at $z = 0.53$. This splits the CMASS sample approximately equally.

We perform analyses with $k_{\max} = 0.15$ and 0.2 .¹¹ We consider $k_{\max} = 0.15$ the conservative baseline choice but $k_{\max} = 0.2$ is still expected to be reliably modeled by `FastPM` as well as the halo model [e.g., 469]. We do not use power spectra on scales larger than $k_{\min} = 0.01$ [487]. Since our theoretical model is simulation based, we do not deconvolve the survey window function. This means that there is a small level of contamination by Fourier modes outside the k -range considered, but we assess this effect to be negligible.

In our baseline analysis we only use the monopole P_0^{gg} of the galaxy auto power spectrum multipoles. This choice was made based on the limited information content of the quadrupole (from EFTofLSS posteriors) and with the aim of simplicity. We discuss the effect of including the quadrupole below. For P^{vg} we use both the monopole and the quadrupole.

It is worth noting that we opt to use reciprocal space void-galaxy cross power spectra P^{vg} instead of the more popular configuration space correlation function. The correlation function has the primary advantage that one can rescale its argument on a void-by-void basis by the respective radius and thus sharpen the resulting void profile (this is also possible in reciprocal space but computationally expensive, future work could explore this point). We believe, however, that the mixing of Fourier modes in the correlation function could lead to problems with approximate solvers like `FastPM`

¹¹For brevity, we implicitly take all wavenumbers in units of $h\text{Mpc}^{-1}$.

whose domain of validity is better localized in reciprocal space. In order to optimize signal-to-noise, we consider P^{vg} computed with three different choices of minimum void radius, 30, 40, and $50 h^{-1}\text{Mpc}$.

An illustration of the data vector is given in Fig. 7.2. We emphasize that the “best-fit” curves shown there are simply picked from the available simulations, implying that the χ^2 values and residuals cannot be interpreted in the usual way.

7.3.2 Implicit likelihood inference

As discussed in the introduction, the standard emulator-based approach is difficult in 17 dimensions. The main reason is that the training objective for an emulator does not directly map to the ultimate goal of accurate posteriors, implying that the optimum needs to be very sharp (which requires many simulations).

Combined with the unknown likelihood function, we believe implicit likelihood inference (ILI) to be the appropriate tool for our task.

We opt for neural ratio estimation (NRE) [488–492] which recasts inference as a classification problem. The choice of an amortized instead of a sequential method¹² was made based on the hierarchical structure of our simulations; we then opt for NRE because of its simplicity. In its original and simplest form, NRE works with pairs of parameter vectors θ, θ' drawn from the prior $p(\theta)$. We then consider a data vector x drawn from the likelihood $p(x|\theta)$, where the simulation process described above approximates this draw. A neural network f maps the pairs $(x, \theta), (x, \theta')$ to scalars y, y' . If we now choose a classification loss function $L(y, y')$, e.g., binary cross entropy

$$L(y, y') = -\log(y) - \log(1 - y'), \quad (7.1)$$

¹²Sequential ILI considers the observed data to sequentially narrow down the space in which to run simulations, thus being more efficient usually. On the other hand, amortized ILI does not specialize on the data at hand; this enables important validations like coverage tests (cf. Fig. 7.6).

it is easy to show that the functional optimization problem

$$f^* = \operatorname{argmin}_f \int d\theta d\theta' dx p(\theta)p(\theta')p(x|\theta) L \quad (7.2)$$

has the solution

$$\frac{p(x|\theta)}{p(x)} = \frac{f^*(x, \theta)}{1 - f^*(x, \theta)}. \quad (7.3)$$

In other words, a neural network trained to distinguish between samples from $p(x, \theta)$ and samples from $p(x)p(\theta)$ approximates the likelihood-to-evidence ratio at optimum. Posteriors can then be obtained through usual Monte Carlo Markov Chain sampling which we perform with `emcee` [366, 367]. In practice, this general idea of approximating $p(x|\theta)/p(x)$ through a classifier works better in the multi-class version “NRE-B” [490]. We use the implementation provided in the `sbi` package [493].

In the above, it is actually not necessary for the parameter vectors θ to be drawn independently from the prior $p(\theta)$. In fact, all that is required is that a sum over the simulated parameter vectors approximates the integral in Eq. (7.2). For this reason, it is correct for us to populate each of the 127 cosmo-varied simulations with multiple draws from the HOD prior (~ 230). For each HOD draw we compute 8 augmentations as described below, yielding $\sim 1.7 \times 10^5$ training samples.

The ILI framework allows implicit marginalization over nuisance parameters. This is one of its primary benefits in high dimensional parameter spaces.¹³ In principle, we could take $\theta = \{\sum m_\nu\}$ as one dimensional. In practice, it is likely better to include a subset of the nuisance parameters in θ . This is because we have intuition for the posteriors expected for some nuisance parameters and thus making them explicit allows useful checks. We opt to include $\log M_{\min}$ and $\mu(M_{\min})$ in the parameter vector. For the former we know that the data should provide a constraint considerably tighter

¹³Consider computation of a high-dimensional integral over $f(\mathbf{x})$ given samples $f(\mathbf{x}_i)$. Interpolating $f(\mathbf{x})$ using these samples and then performing quadrature is a more difficult problem than using the Monte Carlo estimator.

than the prior, while for the latter we expect a result close to zero. The posterior on $\sum m_\nu$ is unaffected by this choice of θ and the extra computational cost in training and sampling the neural network is marginal when making two nuisance parameters explicit (however, making all nuisance parameters explicit would complicate the training unnecessarily).

We parameterize the classifier f as a residual neural network. Hyperparameter optimization was performed considering the loss on a validation set of 13 cosmologies (i.e., $\sim 2.4 \times 10^4$ mocks).¹⁴ We converged at a relatively large network with 1.7×10^7 trainable parameters but high dropout rates.

High-dimensional data vectors x are often problematic for ILI, our problem being no exception. This necessitates a compression step before the data vector is passed to the neural network. Since we expect our likelihood to be close to Gaussian/Poissonian, we use the linear score compression MOPED [364] to 17 compressed statistics. Indeed, MOPED is locally optimal both for a pure Gaussian and a pure Poissonian likelihood. We also experimented with the nuisance-hardened generalization [494] to five and ten dimensions, obtaining consistent but slightly wider posteriors. In order to construct the MOPED compression matrix, estimates for the covariance and derivatives of the data vector are required. We construct the covariance matrix from our fiducial mocks using the usual estimator. For the derivatives, we generate $\sim 10^5$ additional mocks (10^3 parameter vectors, each with 96 augmentations) in a small ball around the fiducial model. We then perform linear regression and read off the derivatives. Simple tests indicate that the dependence on parameters is close to linear in the region considered.

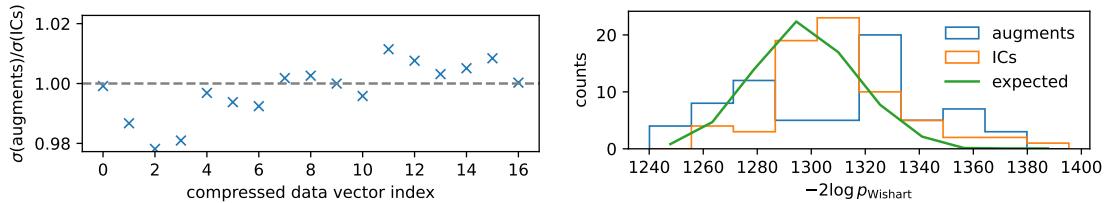


Figure 7.3: Checks for statistical independence of augmentations, in the compressed space. The data labeled ‘‘augments’’ are obtained by marginalizing over augmentations, while those labeled ‘‘ICs’’ are obtained by marginalizing over initial conditions. *Left:* ratio of average standard deviations, difference with unity not exceeding two percent. This check can also be performed with the uncompressed data vector, leading to similar conclusions. *Right:* distributions of log-likelihoods of covariance matrices under the Wishart distribution with our fiducial covariance matrix. The distributions have large overlap, again indicating that the augmentations are very close to statistical independence.

7.3.3 Augmentations

As discussed before, our cosmo-varied simulations share the random seed. This fact ostensibly makes them unsuitable for the ILI approach discussed before since the integral in Eq. (7.2) requires a sampling of initial conditions.

However, as we shall discuss in this section, it is possible to generate many quasi-independent realizations from a single simulation. As mentioned before, we do not require independent identically distributed realizations, so this is in fact enough to approximate Eq. (7.2) with sufficient accuracy.

For a single $2.5 h^{-1}\text{Gpc}$ simulation box populated with galaxies, we can take the product of the following transformations: 2 cuboid remappings, 8 reflections, 6 axis transpositions. This results in 96 augmentations. In principle, many more augmentations can be generated through translations, but we expect these to be more correlated.

The crucial question now is whether these 96 augmentations approximate the distribution over initial conditions. We can answer this question by considering our fiducial simulations which have 69 different random seeds. Given the fiducial parameter

¹⁴It is actually important to separate training and validation sets by cosmologies. Initial trials that mixed simulations exhibited hidden overfitting.

vector, we generate a matrix $D_{\mu a}$ whose elements are data vector-valued and where $\mu = 1 \dots 69$, $a = 1 \dots 96$ index the initial conditions and augmentations, respectively. We can perform statistical tests by computing marginals over μ and a individually or jointly. In order to simplify the statistical interpretation, we restrict a to 69 randomly chosen indices.

In the left panel of Fig. 7.3, we compare the diagonals of covariance matrices in the MOPED compressed space. We see that the standard deviations are almost identical for marginalization over μ and a . This test can also be performed for the uncompressed data vector, yielding consistent results and no systematic differences between different summary statistics or scales.

In the right panel, we perform a test considering the entire content of the covariance matrices. We construct the Wishart distribution given the covariance matrix C_{joint} obtained by marginalizing over μ and a jointly and then compute the log-likelihood of the individually marginalized covariance matrices. If these covariance matrices were drawn from the Wishart distribution sourced by C_{joint} , their log-likelihoods would be distributed as indicated by the green line. We see that the distributions are somewhat different but still have large overlap. In conclusion, the 96 augmentations reproduce the distribution over initial conditions reasonably well. Since $96 \gg 1$, we expect the augmentations to provide a good approximation to the integral in Eq. (7.2).

Why does this augmentation procedure work? First, our simulation boxes are about $5.7 \times$ larger than the survey volume. Second, the augmentations alter the redshift direction. Third, galaxies and the survey mask interact. Fourth, galaxies are captured at different times so their peculiar motions alter their real space positions. All these points need to be seen relative to the specific survey and simulation configuration; the described augmentation procedure is certainly not expected to work universally.

7.4 Results

In this section, we first present our main posteriors on $\sum m_\nu$ from the CMASS NGC data, taking various combinations of the summary statistics VSF N_v , void-galaxy cross power spectrum P^{vg} , and galaxy auto power spectrum P^{gg} .

We present most of our posteriors in their cumulative form. This is because at the current level of precision, no neutrino mass detection is expected and upper bounds are the main objective. The cumulative posterior is the most direct visualization of upper bounds. In all plots we include a diagonal dashed line indicating the prior.

In the following, we will occasionally compare with results obtained with the EFTofLSS [487, 495, 496]. The EFTofLSS allows for the analysis of the full-shape galaxy auto power spectrum (as well as other statistics we will not consider here). We use the window-less full-shape likelihood [497, 498, 455]¹⁵ and the CLASS-PT code [499].¹⁶ We restrict the data included in the likelihood to the NGC high- z sample, approximately equal to the data we use for our analysis. Furthermore, we impose the same Λ CDM prior while keeping the nuisance parameter priors equal to those implemented in the public likelihood code. In any comparison with our results we use identical k_{\max} . Likewise, we usually only use the monopole P_0^{gg} , consistent with our simulation-based analysis. The likelihood part termed “Alcock-Paczynski” in the EFTofLSS likelihood is included, since our method also effectively includes this term. On the other hand, we do not include the BAO reconstruction or real space likelihoods.

We emphasize that a comparison between EFTofLSS and HOD methods is beyond the scope of this work. Therefore, we will use the EFTofLSS posteriors to provide intuition, show that our posteriors are at least qualitatively reasonable, and for an interesting observation about the quadrupole P_2^{gg} later on.

¹⁵https://github.com/oliverphilcox/full_shape_likelihoods

¹⁶<https://github.com/michalychforever/CLASS-PT>

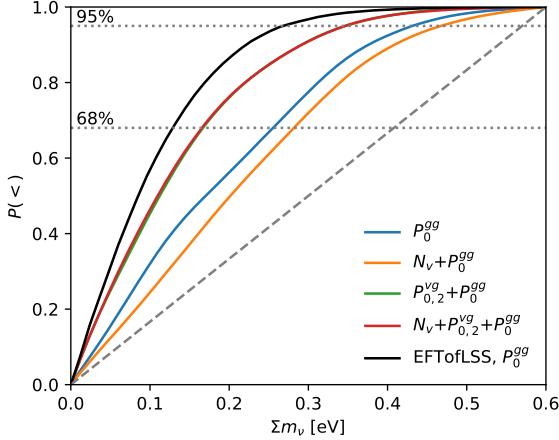


Figure 7.4: Cumulative posteriors on $\sum m_\nu$ from different data vector combinations, with our baseline choice of $k_{\max} = 0.15$. The addition of void-galaxy cross power spectra improves the constraint compared to the galaxy auto power spectrum, whereas the VSF has negligible impact.

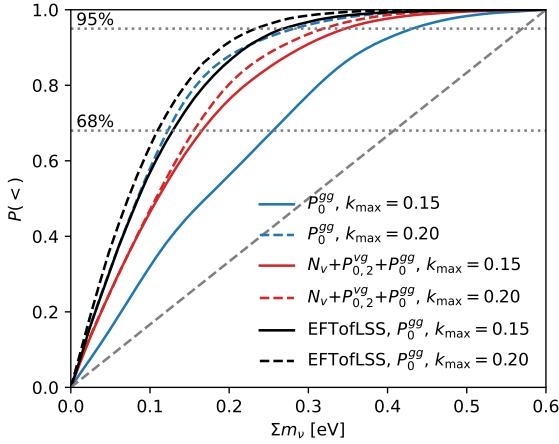


Figure 7.5: Effect of increasing k_{\max} to 0.2. The posterior shrinks, as expected, but remains broader than the EFTofLSS result. The impact of adding voids to the data vector is now reversed.

7.4.1 Neutrino mass posterior

In Fig. 7.4, we show the baseline posterior on $\sum m_\nu$, with $k_{\max} = 0.15$. The galaxy auto power spectrum gives a 95 % credible interval constraint of $\sum m_\nu < 0.43$ eV. Upon inclusion of the VSF, the posterior broadens somewhat. Including the void-galaxy cross power spectrum tightens the posterior to $\sum m_\nu < 0.35$ eV, a $\sim 20\%$ improvement. Further adding the VSF does not lead to any appreciable change. Posteriors are generally wider than the EFTofLSS result.

In Fig. 7.5, we show a similar set of posteriors obtained with $k_{\max} = 0.2$. We believe that our simulated model should still have a high level of fidelity at these somewhat smaller scales. We observe that including smaller scales tightens the posterior, as expected. However, adding void statistics to P^{gg} now slightly broadens the posterior.

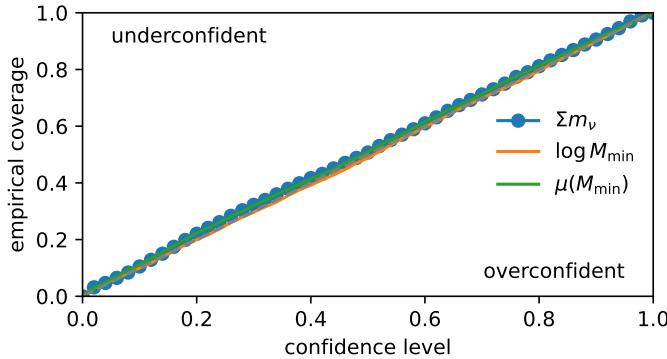


Figure 7.6: Coverage (q-q) plot, demonstrating that when performing inference on mocks the posteriors are well calibrated.

Most of the remainder of this section will be devoted to better understanding the observations from Figs. 7.4, 7.5.

7.4.2 Validation

Any simulation-based, and especially implicit-likelihood, inference necessitates rigorous validation of the simulated model, the likelihood approximation, and the resulting posteriors. In this section, we present three tests verifying different aspects of our pipeline.

First, in Fig. 7.6, we present the usual coverage (or q-q) plot [500]. For this diagnostic, we perform inference on mocks drawn from the prior; in particular, we use the $\sim 2.4 \times 10^4$ validation mocks discussed before. We use the $N_v + P^{vg} + P^{gg}$ likelihood with $k_{\max} = 0.15$. For each resulting chain, we compute the marginal distributions of the explicit parameters and then the confidence level at which the true input parameter is located. In Fig. 7.6, we show the cumulative histograms of these confidence levels. If the posterior is well-calibrated, these CDFs should coincide with the diagonal. As can be seen, for all parameters considered this is the case. This diagnostic is a powerful internal consistency check and verifies that the neural network is well-trained.

Second, in Fig. 7.7, we show an interesting observation concerning the galaxy auto power spectrum quadrupole P_2^{gg} . As discussed before, this summary statistic

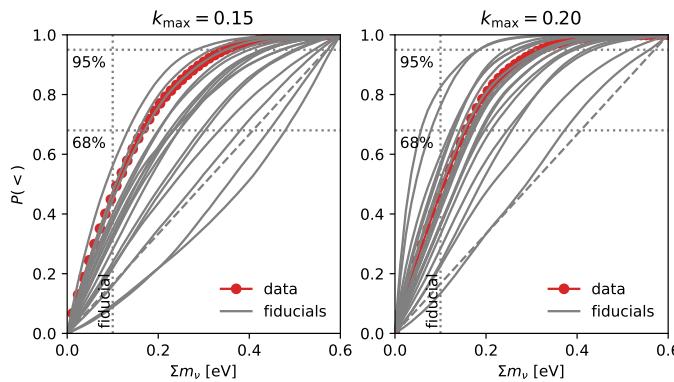
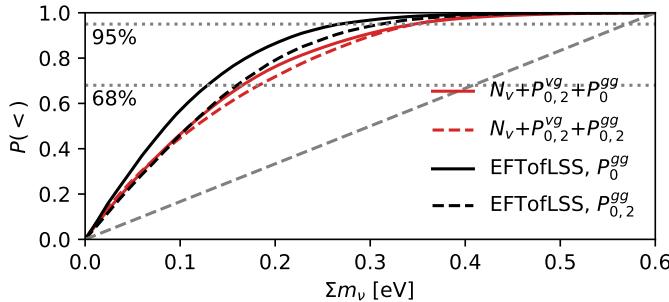


Figure 7.7: Effect of adding the quadrupole P_2^{gg} to the data vector. Consistent with the EFTofLSS result, the $\sum m_\nu$ posterior broadens slightly. This effect is likely a statistical fluctuation. However, it provides a useful cross-check for how well our simulations model redshift space distortions.

Figure 7.8: Comparison of the data posterior with posteriors obtained on fiducial mocks, using the $N_v + P_{0,2}^{vg} + P_0^{gg}$ data vector. Note that the HOD used in generating these mocks is somewhat different from the best-fit. While the data posterior at $k_{\max} = 0.15$ is slightly unusual, at $k_{\max} = 0.20$ it becomes more typical (compared to the distribution of mock posteriors).

has limited constraining power and we do not use it for our main posteriors. As can be seen in Fig. 7.7, adding the quadrupole to the data vector slightly broadens the posterior. This happens consistently in our analysis and in the EFTofLSS. We believe that this observation increases confidence in the validity of our simulation model, in particular the modeling of redshift space distortions.

Third, in Fig. 7.8 we compare the data posteriors with posteriors obtained by running inference on randomly chosen mocks generated at the fiducial point. We remind the reader that the fiducial HOD is rather far from best-fit which somewhat complicates the interpretation. We observe that at $k_{\max} = 0.15$ the data posterior is tighter than most of the mock ones. If the cosmological simulations were to blame for this, the naive expectation would be for the discrepancy to become more severe as smaller scales are included in the analysis. However, this appears not to be the case: at $k_{\max} = 0.20$ the data posterior becomes more typical. We conclude that even though we observe hints of differences between data and simulations, the evidence

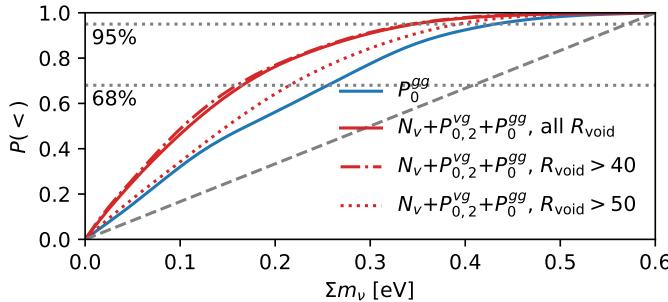


Figure 7.9: Effect of considering voids above different radius cuts. Voids with radii between 40 and $50 h^{-1}\text{Mpc}$ contribute the majority of the observed tightening of the posterior relative to the P^{gg} -only result.

is not conclusive and the data posterior could well be consistent with the observed distribution. It should also be noted that the real $\sum m_\nu$ may be less than the choice 0.1 eV with which the fiducials were run, potentially leading to a tighter data posterior.

7.4.3 Broadening of posteriors

One peculiar observation is that inclusion of void statistics can broaden the posterior on $\sum m_\nu$. We do not fully understand this phenomenon and can only provide some suggestive results. These are more comprehensively described in appendix F.2; here we only provide a summary.

We observe similar broadening on fiducial mocks and thus propose that we are in fact observing a generic phenomenon. Therefore, we suggest that void statistics are most effective at constraining the neutrino mass sum from below. A further test using artificially enlarged volumes supports this theory.

For a potential physical explanation, we consider the free streaming length. At $z = 0.5$, $\lambda_{\text{fs}} = 90 h^{-1}\text{Mpc}$ ($0.3\text{ eV}/\sum m_\nu$) for degenerate masses. This length scale is comparable to the diameter of the voids that seem to contribute most (c.f. Sec. 7.4.4). Thus, it may be that $\sum m_\nu$ at the upper end of the posterior is “invisible” to voids. However, we identify voids using tracers of small-scale fluctuations, so the full picture is much more complicated and could be a subject for further study.

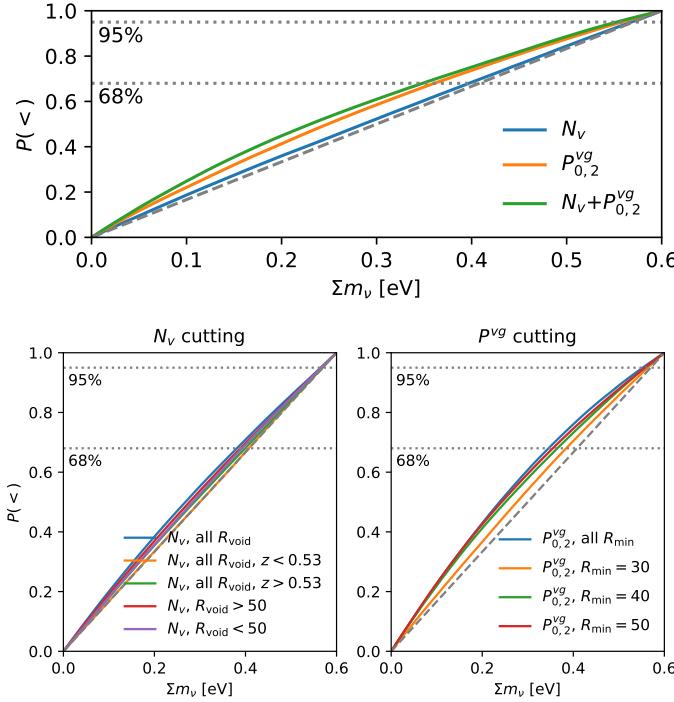


Figure 7.10: Different void-only statistics combinations. *Upper panel*: while all posteriors are close to the prior, it appears that P^{vg} carries somewhat more signal than N_ν , and the combined signal is strongest. *Lower panel*: consistent with Fig. 7.9, it appears that larger voids carry more neutrino mass signal, although the posteriors are quite close to the prior.

7.4.4 Void radius

In Fig. 7.9, we show posteriors obtained with the $N_\nu + P^{vg} + P^{gg}$ likelihood, concentrating on void size. Cuts on effective radius are performed both in the VSF and P^{vg} parts of the data vector. We observe that the posteriors are almost identical regardless of whether we cut at 30 (the baseline analysis) or $40 h^{-1}\text{Mpc}$. On the other hand, further increasing the minimum radius to $50 h^{-1}\text{Mpc}$ removes much of the effect of voids on the posterior. Fig. 7.9 indicates that at least for the present analysis voids with effective radii between 40 and $50 h^{-1}\text{Mpc}$ are the most constraining. Smaller voids might be contaminated by spurious Poisson voids and perhaps also due to their shallower density profile less affected by neutrinos. Larger voids presumably suffer from their low abundance.

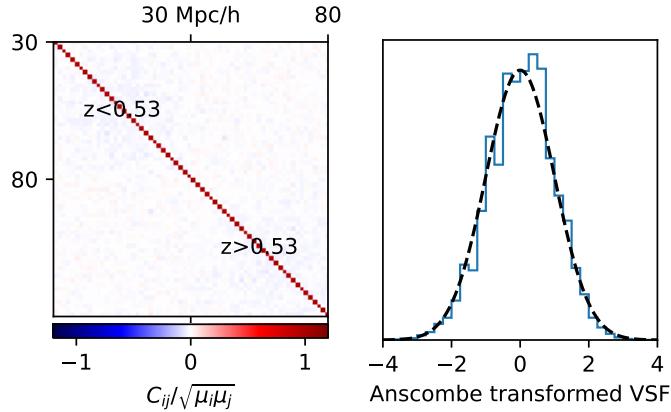


Figure 7.11: Checks for Poissonian nature of the VSF. In the left panel, we show a rescaled covariance matrix obtained from our fiducial simulations of the VSF part of the data vector. For an exactly Poissonian distribution, this matrix would be the identity. In the right panel, we show the distribution of Anscombe transformed VSF counts. Overplotted is a standard normal. Both tests indicate a distribution that is very close to Poissonian.

In Fig. 7.10 we show posteriors obtained from void statistics only. We show them mostly for completeness; in the present analysis these are entirely prior dominated. However, even in this plot we see the previously mentioned observation that larger voids appear to carry more signal.

7.4.5 Poissonian void size function

As a final point of this section, we substantiate the previous claim that the VSF is very close to Poissonian distributed. While this seems to be a natural assumption, void exclusion makes it non-trivial. Indeed, previous works have assumed Poisson likelihoods [393, 418]; our simulations enable us to check this assumption.

In Fig. 7.11, we show two checks performed with our fiducial mocks. The left panel shows the covariance matrix divided by the outer square of Poissonian standard deviations; the result is close to the identity. The right panel shows a check using the variance-stabilizing Anscombe transform [501]. For each mock data vector $c^{(\alpha)}$ and

bin i , we compute the transformed VSF count

$$\tilde{c}_i^{(\alpha)} = 2 \left(\sqrt{c_i^{(\alpha)} + \frac{3}{8}} - \sqrt{\langle c_i \rangle + \frac{3}{8}} \right) + \frac{1}{4\sqrt{\langle c_i \rangle}}. \quad (7.4)$$

In the limit of large counts the distribution of these transformed counts converges to the standard normal if the counts themselves are Poissonian. As can be seen, the agreement with the standard normal is quite good indeed. These tests demonstrate that deviations from Poissonian distribution are small for the VSF, at least for the choice of binning considered here.

7.5 Conclusions

We have performed inference on galaxy clustering in the BOSS CMASS northern sample, combining the void size function, the void-galaxy cross power spectrum, and the galaxy auto power spectrum. Our primary target was the neutrino mass sum, $\sum m_\nu$; thus, we imposed a tight prior on Λ CDM informed by primary CMB data.

We argued that analytic models for the considered void statistics are not mature enough and unsuitable for our specific problem, necessitating a simulation-based approach. To this end, we ran approximate gravity-only simulations and populated them with galaxies using an expressive halo occupation distribution. Several factors motivated the use of implicit likelihood inference.

In our baseline analysis, we find $\sum m_\nu < 0.43$ eV from the galaxy auto power spectrum alone, and $\sum m_\nu < 0.35$ eV with the void statistics included (95 % credible interval). We performed several tests to confirm statistical and systematic validity of our likelihood approximation.

We performed a short investigation of the impact of voids on the neutrino mass posterior. It appears that the void statistics may be most effective in constraining

$\sum m_\nu$ from below. This result would imply that future analyses aiming at *measuring* $\sum m_\nu$ may benefit from including void statistics.

Our results suggest that larger voids with effective radii $> 40 h^{-1}\text{Mpc}$ carry most of the signal despite their lower abundance. This has interesting implications for future analyses, since voids of this size should be detectable in photometric catalogs with relatively low redshift error, such as the one expected for Rubin/LSST [329]. Of course, spectroscopic surveys such as DESI [502], Euclid [330], SPHEREx [503], PFS [504], and Roman [331] will continue to be cornerstones of void science. The trade-off between volume, galaxy number density, and redshift precision warrants further investigation.

We also demonstrate that the void size function is very close to Poisson distributed, a feature that had been assumed in previous analyses but never explicitly confirmed.

Future work could improve upon our analysis in multiple ways. First, the cosmovaried simulations should be run with different random seeds (we decided for a fixed seed in anticipation of an emulator-based analysis which ultimately turned out to be very difficult). Second, it may be beneficial to normalize the void-galaxy cross power spectrum by void number. Although in principle this would contain the same information as our data vector once the VSF is included, the necessary transformation is non-linear and thus potentially invisible to our data compression. Third, the HOD modeling could be improved. Some of our priors may not be optimal, and our $n(z)$ downsampling is simplistic. The CMASS sample's completeness is quite well known and could be used to put a prior on the downsampling. Fourth, it turns out that the cosmological simulations did not dominate compute cost. It may therefore be economical to increase accuracy in `FastPM` or switch to a different solver.

Our results point toward a complicated picture with regard to the relationship between massive neutrinos and voids. Future data sets, both spectroscopic as well as

photometric, promise to bring tight cosmological constraints from void science, since it scales well with number.

Acknowledgements. We thank Sofia Contarini, Adrian Bayer, Jia Liu, Jo Dunkley, Masahiro Takada for useful discussions. We thank Oliver Philcox for explaining the EFTofLSS likelihood. The work of LT is supported by the NSF grant AST 2108078. The authors are pleased to acknowledge that the work reported on in this paper was substantially performed using the Princeton Research Computing resources at Princeton University which is a consortium of groups led by the Princeton Institute for Computational Science and Engineering (PICSciE) and Office of Information Technology’s Research Computing.

Bibliography

- [1] Planck Collaboration, N. Aghanim, M. Arnaud, M. Ashdown, et al. Planck 2015 results. XI. CMB power spectra, likelihoods, and robustness of parameters. *Astron. Astrophys.*, 594:A11, September 2016.
- [2] Leander Thiele, Yilun Guan, J. Colin Hill, Arthur Kosowsky, et al. Can small-scale baryon inhomogeneities resolve the Hubble tension? An investigation with ACT DR4. *Phys. Rev. D*, 104(6):063535, September 2021.
- [3] Adam G. Riess, Stefano Casertano, Wenlong Yuan, J. Bradley Bowers, et al. Cosmic Distances Calibrated to 1% Precision with Gaia EDR3 Parallaxes and Hubble Space Telescope Photometry of 75 Milky Way Cepheids Confirm Tension with Λ CDM. *Astrophys. J. Lett.*, 908(1):L6, February 2021.
- [4] Wendy L. Freedman, Barry F. Madore, Taylor Hoyt, In Sung Jang, et al. Calibration of the Tip of the Red Giant Branch. *Astrophys. J.*, 891(1):57, March 2020.
- [5] M. J. Reid, J. A. Braatz, J. J. Condon, L. J. Greenhill, et al. The Megamaser Cosmology Project. I. Very Long Baseline Interferometric Observations of UGC 3789. *Astrophys. J.*, 695(1):287–291, April 2009.
- [6] Kenneth C. Wong, Sherry H. Suyu, Geoff C. F. Chen, Cristian E. Rusu, et al. H0LiCOW – XIII. A 2.4 per cent measurement of H_0 from lensed quasars: 5.3σ tension between early- and late-Universe probes. *Mon. Not. Roy. Astron. Soc.*, 498(1):1420–1439, October 2020.
- [7] S. Birrer, A. J. Shajib, A. Galan, M. Millon, et al. TDCOSMO. IV. Hierarchical time-delay cosmography – joint inference of the Hubble constant and galaxy density profiles. *Astron. Astrophys.*, 643:A165, November 2020.
- [8] Planck Collaboration. Planck 2018 results. VI. Cosmological parameters. *Astron. Astrophys.*, 641:A6, September 2020.
- [9] ACT Collaboration, Simone Aiola, Erminia Calabrese, Loïc Maurin, et al. The Atacama Cosmology Telescope: DR4 maps and cosmological parameters. *J. Cosmol. Astropart. Phys.*, 2020(12):047, December 2020.

- [10] C. L. Bennett, D. Larson, J. L. Weiland, N. Jarosik, et al. Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Final Maps and Results. *Astrophys. J. Suppl.*, 208(2):20, October 2013.
- [11] Oliver H. E. Philcox, Mikhail M. Ivanov, Marko Simonović, and Matias Zaldarriaga. Combining full-shape and BAO analyses of galaxy power spectra: a 1.6% CMB-independent constraint on H_0 . *J. Cosmol. Astropart. Phys.*, 2020(5):032, May 2020.
- [12] L. Knox and M. Millea. Hubble constant hunter's guide. *Phys. Rev. D*, 101(4):043533, February 2020.
- [13] Tanvi Karwal and Marc Kamionkowski. Dark energy at early times, the Hubble parameter, and the string axiverse. *Phys. Rev. D*, 94(10):103523, November 2016.
- [14] Vivian Poulin, Tristan L. Smith, Tanvi Karwal, and Marc Kamionkowski. Early Dark Energy can Resolve the Hubble Tension. *Phys. Rev. Lett.*, 122(22):221301, June 2019.
- [15] Prateek Agrawal, Francis-Yan Cyr-Racine, David Pinner, and Lisa Randall. Rock 'n' Roll Solutions to the Hubble Tension. *arXiv e-prints*, page arXiv:1904.01016, April 2019.
- [16] Meng-Xiang Lin, Giampaolo Benevento, Wayne Hu, and Marco Raveri. Acoustic dark energy: Potential conversion of the Hubble tension. *Phys. Rev. D*, 100(6):063542, September 2019.
- [17] Jeremy Sakstein and Mark Trodden. Early Dark Energy from Massive Neutrinos as a Natural Resolution of the Hubble Tension. *Phys. Rev. Lett.*, 124(16):161301, April 2020.
- [18] Gen Ye and Yun-Song Piao. Is the Hubble tension a hint of AdS phase around recombination? *Phys. Rev. D*, 101(8):083507, April 2020.
- [19] J. Colin Hill, Evan McDonough, Michael W. Toomey, and Stephon Alexander. Early dark energy does not restore cosmological concordance. *Phys. Rev. D*, 102(4):043507, August 2020.
- [20] Mikhail M. Ivanov, Evan McDonough, J. Colin Hill, Marko Simonović, et al. Constraining early dark energy with large-scale structure. *Phys. Rev. D*, 102(10):103502, November 2020.
- [21] Guido D'Amico, Leonardo Senatore, Pierre Zhang, and Henry Zheng. The Hubble Tension in Light of the Full-Shape Analysis of Large-Scale Structure Data. *arXiv e-prints*, page arXiv:2006.12420, June 2020.

- [22] Francesco D'Eramo, Ricardo Z. Ferreira, Alessio Notari, and José Luis Bernal. Hot axions and the H_0 tension. *J. Cosmol. Astropart. Phys.*, 2018(11):014, November 2018.
- [23] Christina D. Kreisch, Francis-Yan Cyr-Racine, and Olivier Doré. Neutrino puzzle: Anomalies, interactions, and cosmological tensions. *Phys. Rev. D*, 101(12):123505, June 2020.
- [24] Anirban Das and Subhajit Ghosh. Flavor-specific Interaction Favors Strong Neutrino Self-coupling in the Early Universe. *arXiv e-prints*, page arXiv:2011.12315, November 2020.
- [25] Thejs Brinckmann, Jae Hyeok Chang, and Marilena LoVerde. Self-interacting neutrinos, the Hubble parameter tension, and the Cosmic Microwave Background. *arXiv e-prints*, page arXiv:2012.11830, December 2020.
- [26] Shouvik Roy Choudhury, Steen Hannestad, and Thomas Tram. Updated constraints on massive neutrino self-interactions from cosmology in light of the H_0 tension. *arXiv e-prints*, page arXiv:2012.07519, December 2020.
- [27] Qing-Guo Huang and Ke Wang. How the dark energy can reconcile Planck with local determination of the Hubble constant. *European Physical Journal C*, 76(9):506, September 2016.
- [28] Janina Renk, Miguel Zumalacárregui, Francesco Montanari, and Alexandre Barreira. Galileon gravity in light of ISW, CMB, BAO and H_0 data. *J. Cosmol. Astropart. Phys.*, 2017(10):020, October 2017.
- [29] Gong-Bo Zhao, Marco Raveri, Levon Pogosian, Yuting Wang, et al. Dynamical dark energy in light of the latest observations. *Nature Astronomy*, 1:627–632, August 2017.
- [30] Rafael C. Nunes. Structure formation in $f(T)$ gravity and a solution for H_0 tension. *J. Cosmol. Astropart. Phys.*, 2018(5):052, May 2018.
- [31] Eleonora Di Valentino, Eric V. Linder, and Alessandro Melchiorri. Vacuum phase transition solves the H_0 tension. *Phys. Rev. D*, 97(4):043528, February 2018.
- [32] Yuting Wang, Levon Pogosian, Gong-Bo Zhao, and Alex Zucca. Evolution of Dark Energy Reconstructed from the Latest Observations. *Astrophys. J. Lett.*, 869(1):L8, December 2018.
- [33] Nima Khosravi, Shant Baghram, Niayesh Afshordi, and Natacha Altamirano. H_0 tension as a hint for a transition in gravitational theory. *Phys. Rev. D*, 99(10):103526, May 2019.
- [34] Marco Raveri. Reconstructing gravity on cosmological scales. *Phys. Rev. D*, 101(8):083524, April 2020.

- [35] Abdolali Banihashemi, Nima Khosravi, and Amir H. Shirazi. Phase transition in the dark sector as a proposal to lessen cosmological tensions. *Phys. Rev. D*, 101(12):123521, June 2020.
- [36] Yi-Fu Cai, Martiros Khurshudyan, and Emmanuel N. Saridakis. Model-independent Reconstruction of f(T) Gravity from Gaussian Processes. *Astrophys. J.*, 888(2):62, January 2020.
- [37] Eleonora Di Valentino, Alessandro Melchiorri, Olga Mena, and Sunny Vagnozzi. Nonminimal dark sector physics and cosmological tensions. *Phys. Rev. D*, 101(6):063502, March 2020.
- [38] Maria Giovanna Dainotti, Biagio De Simone, Tiziano Schiavone, Giovanni Montani, et al. On the Hubble constant tension in the SNe Ia Pantheon sample. *arXiv e-prints*, page arXiv:2103.02117, March 2021.
- [39] Harry Desmond, Bhuvnesh Jain, and Jeremy Sakstein. Local resolution of the Hubble tension: The impact of screened fifth forces on the cosmic distance ladder. *Phys. Rev. D*, 100(4):043537, August 2019.
- [40] George Efstathiou. To H0 or not to H0? *arXiv e-prints*, page arXiv:2103.08723, March 2021.
- [41] Karsten Jedamzik, Levon Pogosian, and Gong-Bo Zhao. Why reducing the cosmic sound horizon can not fully resolve the Hubble tension. *arXiv e-prints*, page arXiv:2010.04158, October 2020.
- [42] P. J. E. Peebles. *Principles of Physical Cosmology*. 1976.
- [43] Karsten Jedamzik and Tom Abel. Weak Primordial Magnetic Fields and Anisotropies in the Cosmic Microwave Background Radiation. *arXiv e-prints*, page arXiv:1108.2517, August 2011.
- [44] Robi Banerjee and Karsten Jedamzik. Evolution of cosmic magnetic fields: From the very early Universe, to recombination, to the present. *Phys. Rev. D*, 70(12):123003, December 2004.
- [45] Karsten Jedamzik and Andrey Saveliev. Stringent Limit on Primordial Magnetic Fields from the Cosmic Microwave Background Radiation. *Phys. Rev. Lett.*, 123(2):021301, July 2019.
- [46] Karsten Jedamzik and Levon Pogosian. Relieving the Hubble Tension with Primordial Magnetic Fields. *Phys. Rev. Lett.*, 125(18):181302, October 2020.
- [47] MiaoXin Liu, ZhiQi Huang, XiaoLin Luo, HaiTao Miao, et al. Can non-standard recombination resolve the Hubble tension? *Science China Physics, Mechanics, and Astronomy*, 63(9):290405, May 2020.

- [48] ACT Collaboration, Steve K. Choi, Matthew Hasselfield, Shuay-Pwu Patty Ho, et al. The Atacama Cosmology Telescope: a measurement of the Cosmic Microwave Background power spectra at 98 and 150 GHz. *J. Cosmol. Astropart. Phys.*, 2020(12):045, 2020.
- [49] Antony Lewis, Anthony Challinor, and Anthony Lasenby. Efficient Computation of Cosmic Microwave Background Anisotropies in Closed Friedmann-Robertson-Walker Models. *Astrophys. J.*, 538(2):473–476, August 2000.
- [50] S. Seager, D. D. Sasselov, and D. Scott. A New Calculation of the Recombination Epoch. *Astrophys. J. Lett.*, 523(1):L1–L5, September 1999.
- [51] Sara Seager, Dimitar D. Sasselov, and Douglas Scott. How Exactly Did the Universe Become Neutral? *Astrophys. J. Suppl.*, 128(2):407–430, June 2000.
- [52] Wan Yan Wong, Adam Moss, and Douglas Scott. How well do we understand cosmological recombination? *Mon. Not. Roy. Astron. Soc.*, 386(2):1023–1028, May 2008.
- [53] Jesus Torrado and Antony Lewis. Cobaya: Code for Bayesian Analysis of hierarchical physical models. *arXiv e-prints*, page arXiv:2005.05290, May 2020.
- [54] Jesús Torrado and Antony Lewis. Cobaya: Bayesian analysis in cosmology, October 2019.
- [55] Julien Lesgourgues. The Cosmic Linear Anisotropy Solving System (CLASS) I: Overview. *arXiv e-prints*, page arXiv:1104.2932, April 2011.
- [56] Diego Blas, Julien Lesgourgues, and Thomas Tram. The Cosmic Linear Anisotropy Solving System (CLASS). Part II: Approximation schemes. *J. Cosmol. Astropart. Phys.*, 2011(7):034, July 2011.
- [57] Yacine Ali-Haïmoud and Christopher M. Hirata. HyRec: A fast and highly accurate primordial hydrogen and helium recombination code. *Phys. Rev. D*, 83(4):043513, February 2011.
- [58] Thejs Brinckmann and Julien Lesgourgues. MontePython 3: boosted MCMC sampler and other features. 2018.
- [59] Benjamin Audren, Julien Lesgourgues, Karim Benabed, and Simon Prunet. Conservative Constraints on Early Cosmology: an illustration of the Monte Python cosmological parameter inference code. *JCAP*, 1302:001, 2013.
- [60] Fiona McCarthy, J. Colin Hill, and Mathew S. Madhavacheril. Baryonic feedback biases on fundamental physics from lensed CMB power spectra. *arXiv e-prints*, page arXiv:2103.05582, March 2021.
- [61] Planck Collaboration. Planck 2018 results. V. CMB power spectra and likelihoods. *Astron. Astrophys.*, 641:A5, September 2020.

- [62] M. J. Reid, D. W. Pesce, and A. G. Riess. An Improved Distance to NGC 4258 and Its Implications for the Hubble Constant. *Astrophys. J. Lett.*, 886(2):L27, December 2019.
- [63] Ashley J. Ross, Lado Samushia, Cullan Howlett, Will J. Percival, et al. The clustering of the SDSS DR7 main Galaxy sample - I. A 4 per cent distance measure at $z = 0.15$. *Mon. Not. Roy. Astron. Soc.*, 449(1):835–847, May 2015.
- [64] SDSS-III Collaboration. The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: baryon acoustic oscillations in the Data Release 9 spectroscopic galaxy sample. *Mon. Not. Roy. Astron. Soc.*, 427(4):3435–3467, December 2012.
- [65] Florian Beutler, Chris Blake, Matthew Colless, D. Heath Jones, et al. The 6dF Galaxy Survey: baryon acoustic oscillations and the local Hubble constant. *Mon. Not. Roy. Astron. Soc.*, 416(4):3017–3032, October 2011.
- [66] Matías Zaldarriaga and Diego D. Harari. Analytic approach to the polarization of the cosmic microwave background in flat and open universes. *Phys. Rev. D*, 52(6):3276–3287, September 1995.
- [67] SPT-3G Collaboration, B. A. Benson, et al. SPT-3G: a next-generation cosmic microwave background polarization experiment on the South Pole telescope. In Wayne S. Holland and Jonas Zmuidzinas, editors, *Millimeter, Submillimeter, and Far-Infrared Detectors and Instrumentation for Astronomy VII*, volume 9153 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 91531P, July 2014.
- [68] Simons Observatory Collaboration. The Simons Observatory: science goals and forecasts. *J. Cosmol. Astropart. Phys.*, 2019(2):056, February 2019.
- [69] CMB-S4 Collaboration. CMB-S4 Science Case, Reference Design, and Project Plan. *arXiv e-prints*, page arXiv:1907.04473, July 2019.
- [70] Leander Thiele, Digvijay Wadekar, J. Colin Hill, Nicholas Battaglia, et al. Percent-level constraints on baryonic feedback with spectral distortion measurements. *Phys. Rev. D*, 105(8):083505, April 2022.
- [71] R. A. Sunyaev and R. Khatri. Unavoidable CMB Spectral Features and Blackbody Photosphere of Our Universe. *IJMPD*, 22:30014, June 2013.
- [72] J. Chluba. Which spectral distortions does Λ CDM actually predict? *Mon. Not. Roy. Astron. Soc.*, 460:227–239, July 2016.
- [73] Ya. B. Zeldovich and R. A. Sunyaev. The Interaction of Matter and Radiation in a Hot-Model Universe. *Astrophys. Space Sci.*, 4:301–316, July 1969.
- [74] R. A. Sunyaev and Ya. B. Zeldovich. Small-Scale Fluctuations of Relic Radiation. *Astrophys. Space Sci.*, 7:3–19, April 1970.

- [75] J. E. Carlstrom, G. P. Holder, and E. D. Reese. Cosmology with the Sunyaev-Zel'dovich Effect. *Annu. Rev. Astron. Astrophys.*, 40:643–680, 2002.
- [76] T. Mroczkowski, D. Nagai, K. Basu, J. Chluba, et al. Astrophysics with the Spatially and Spectrally Resolved Sunyaev-Zeldovich Effects. A Millimetre/Sub-millimetre Probe of the Warm and Hot Universe. *Space Science Reviews*, 215:17, February 2019.
- [77] Francisco Villaescusa-Navarro, Daniel Anglés-Alcázar, Shy Genel, David N. Spergel, et al. The CAMELS Project: Cosmology and Astrophysics with Machine-learning Simulations. *Astrophys. J.*, 915(1):71, July 2021.
- [78] Maximilian H. Abitbol, Jens Chluba, J. Colin Hill, and Bradley R. Johnson. Prospects for measuring cosmic microwave background spectral distortions in the presence of foregrounds. *Mon. Not. Roy. Astron. Soc.*, 471(1):1126–1140, October 2017.
- [79] S. Y. Sazonov and R. A. Sunyaev. Cosmic Microwave Background Radiation in the Direction of a Moving Cluster of Galaxies with Hot Gas: Relativistic Corrections. *Astrophys. J.*, 508(1):1–5, November 1998.
- [80] Anthony Challinor and Anthony Lasenby. Relativistic Corrections to the Sunyaev-Zeldovich Effect. *Astrophys. J.*, 499(1):1–6, May 1998.
- [81] Naoki Itoh, Yasuharu Kohyama, and Satoshi Nozawa. Relativistic Corrections to the Sunyaev-Zeldovich Effect for Clusters of Galaxies. *Astrophys. J.*, 502(1):7–15, July 1998.
- [82] J. Chluba, G. Hütsi, and R. A. Sunyaev. Clusters of galaxies in the microwave band: Influence of the motion of the Solar System. *Astron. Astrophys.*, 434(3):811–817, May 2005.
- [83] S. Nozawa, N. Itoh, Y. Suda, and Y. Ohhata. An improved formula for the relativistic corrections to the kinematical Sunyaev-Zeldovich effect for clusters of galaxies. *Nuovo Cimento B Serie*, 121(5):487–500, May 2006.
- [84] J. Chluba, D. Nagai, S. Sazonov, and K. Nelson. A fast and accurate method for computing the Sunyaev-Zel'dovich signal of hot galaxy clusters. *Mon. Not. Roy. Astron. Soc.*, 426:510–530, October 2012.
- [85] J. Colin Hill, Nick Battaglia, Jens Chluba, Simone Ferraro, et al. Taking the Universe's Temperature with Spectral Distortions of the Cosmic Microwave Background. *Phys. Rev. Lett.*, 115(26):261301, December 2015.
- [86] S. T. Kay, L. C. Powell, A. R. Liddle, and P. A. Thomas. The Sunyaev-Zel'dovich temperature of the intracluster medium. *Mon. Not. Roy. Astron. Soc.*, 386:2110–2114, June 2008.

- [87] Elizabeth Lee, Jens Chluba, Scott T. Kay, and David J. Barnes. Relativistic SZ temperature scaling relations of groups and clusters derived from the BAHAMAS and MACSIS simulations. *Mon. Not. Roy. Astron. Soc.*, 493(3):3274–3292, April 2020.
- [88] K. Jedamzik, V. Katalinić, and A. V. Olinto. Limit on Primordial Small-Scale Magnetic Fields from Cosmic Microwave Background Distortions. *Phys. Rev. Lett.*, 85:700–703, July 2000.
- [89] Kerstin E. Kunze and Eiichiro Komatsu. Constraining primordial magnetic fields with distortions of the black-body spectrum of the cosmic microwave background: pre- and post-decoupling contributions. *J. Cosmol. Astropart. Phys.*, 2014(1):009, January 2014.
- [90] Jens Chluba, D. Paoletti, F. Finelli, and J. A. Rubiño-Martín. Effect of primordial magnetic fields on the ionization history. *Mon. Not. Roy. Astron. Soc.*, 451(2):2244–2250, August 2015.
- [91] S. Sarkar and A. M. Cooper. Cosmological and experimental constraints on the tau neutrino. *Phys. Lett. B*, 148:347–354, November 1984.
- [92] W. Hu and J. Silk. Thermalization constraints and spectral distortions for massive unstable relic particles. *Physical Review Letters*, 70:2661–2664, May 1993.
- [93] J. Chluba. Distinguishing different scenarios of early energy release with spectral distortions of the cosmic microwave background. *Mon. Not. Roy. Astron. Soc.*, 436:2232–2243, December 2013.
- [94] Yacine Ali-Haïmoud, Jens Chluba, and Marc Kamionkowski. Constraints on Dark Matter Interactions with Standard Model Particles from Cosmic Microwave Background Spectral Distortions. *Phys. Rev. Lett.*, 115:071304, Aug 2015.
- [95] Boris Bolliet, Jens Chluba, and Richard Battye. Spectral distortion constraints on photon injection from low-mass decaying particles. *arXiv e-prints*, page arXiv:2012.07292, December 2020.
- [96] Yacine Ali-Haïmoud. Testing dark matter interactions with CMB spectral distortions. *Phys. Rev. D*, 103(4):043541, February 2021.
- [97] D. J. Fixsen, E. S. Cheng, J. M. Gales, J. C. Mather, et al. The Cosmic Microwave Background Spectrum from the Full COBE FIRAS Data Set. *Astrophys. J.*, 473:576, December 1996.
- [98] A. Kogut, D. J. Fixsen, D. T. Chuss, J. Dotson, et al. The Primordial Inflation Explorer (PIXIE): a nulling polarimeter for cosmic microwave background observations. *J. Cosmol. Astropart. Phys.*, 2011(7):025, July 2011.

- [99] A. Kogut, J. Chluba, D. J. Fixsen, S. Meyer, et al. The Primordial Inflation Explorer (PIXIE). In *SPIE Conference Series*, volume 9904 of *Proc.SPIE*, page 99040W, July 2016.
- [100] B. Maffei, M. H. Abitbol, N. Aghanim, J. Aumont, et al. BISOU: a balloon project to measure the CMB spectral distortions. *arXiv e-prints*, page arXiv:2111.00246, October 2021.
- [101] J. Chluba, M. H. Abitbol, N. Aghanim, Y. Ali-Haïmoud, et al. New horizons in cosmology with spectral distortions of the cosmic microwave background. *Experimental Astronomy*, 51(3):1515–1554, June 2021.
- [102] T. Plagge, B. A. Benson, et al. Sunyaev-Zel'dovich Cluster Profiles Measured with the South Pole Telescope. *Astrophys. J.*, 716(2):1118–1135, June 2010.
- [103] Nick Hand, Atacama Cosmology Telescope Collaboration, et al. The Atacama Cosmology Telescope: Detection of Sunyaev-Zel'Dovich Decrement in Groups and Clusters Associated with Luminous Red Galaxies. *Astrophys. J.*, 736(1):39, July 2011.
- [104] Planck Collaboration et al. Planck intermediate results. XI. The gas content of dark matter halos: the Sunyaev-Zeldovich-stellar mass relation for locally brightest galaxies. *Astron. Astrophys.*, 557:A52, September 2013.
- [105] Planck Collaboration, P. A. R. Ade, N. Aghanim, M. Arnaud, et al. Planck intermediate results. V. Pressure profiles of galaxy clusters from the Sunyaev-Zeldovich effect. *Astron. Astrophys.*, 550:A131, February 2013.
- [106] J. Colin Hill, Blake D. Sherwin, Kendrick M. Smith, Atacama Cosmology Telescope Collaboration, et al. The Atacama Cosmology Telescope: A Measurement of the Thermal Sunyaev-Zel'dovich One-Point PDF. *arXiv e-prints*, page arXiv:1411.8004, November 2014.
- [107] J. P. Greco, J. C. Hill, D. N. Spergel, and N. Battaglia. The Stacked Thermal Sunyaev-Zel'dovich Signal of Locally Brightest Galaxies in Planck Full Mission Data: Evidence for Galaxy Feedback? *Astrophys. J.*, 808:151, August 2015.
- [108] Planck Collaboration et al. Planck 2015 results. XXII. A map of the thermal Sunyaev-Zeldovich effect. *Astron. Astrophys.*, 594:A22, September 2016.
- [109] V. Vikram, A. Lidz, and B. Jain. A Measurement of the Galaxy Group-Thermal Sunyaev-Zel'dovich Effect Cross-Correlation Function. *Mon. Not. Roy. Astron. Soc.*, 467:2315–2330, May 2017.
- [110] J. P. Dietrich, S. Bocquet, T. Schrabback, D. Applegate, et al. Sunyaev-Zel'dovich effect and X-ray scaling relations from weak lensing mass calibration of 32 South Pole Telescope selected galaxy clusters. *Mon. Not. Roy. Astron. Soc.*, 483(3):2871–2906, March 2019.

- [111] Mathew S. Madhavacheril, J. Colin Hill, Sigurd Næss, Atacama Cosmology Telescope Collaboration, et al. Atacama Cosmology Telescope: Component-separated maps of CMB temperature and the thermal Sunyaev-Zel'dovich effect. *Phys. Rev. D*, 102(2):023534, July 2020.
- [112] L. E. Bleem, T. M. Crawford, et al. CMB/kSZ and Compton- y Maps from 2500 square degrees of SPT-SZ and Planck Survey Data. *arXiv e-prints*, page arXiv:2102.05033, February 2021.
- [113] Emmanuel Schaan, Simone Ferraro, Stefania Amodeo, Nicholas Battaglia, et al. Atacama Cosmology Telescope: Combined kinematic and thermal Sunyaev-Zel'dovich measurements from BOSS CMASS and LOWZ halos. *Phys. Rev. D*, 103(6):063513, March 2021.
- [114] Stefania Amodeo, Nicholas Battaglia, Emmanuel Schaan, Simone Ferraro, et al. Atacama Cosmology Telescope: Modeling the gas thermodynamics in BOSS CMASS galaxies from kinematic and thermal Sunyaev-Zel'dovich measurements. *Phys. Rev. D*, 103(6):063514, March 2021.
- [115] E. M. Vavagiakis, P. A. Gallardo, V. Calafut, S. Amodeo, et al. The Atacama Cosmology Telescope: Probing the baryon content of SDSS DR15 galaxies with the thermal and kinematic Sunyaev-Zel'dovich effects. *Phys. Rev. D*, 104(4):043503, August 2021.
- [116] Cameron T. Pratt, Zhijie Qu, and Joel N. Bregman. The Resolved Sunyaev-Zel'dovich Profiles of Nearby Galaxy Groups. *Astrophys. J.*, 920(2):104, October 2021.
- [117] G. Hurier and C. Tchernin. Mapping the temperature of the intra-cluster medium with the thermal Sunyaev-Zel'dovich effect. *Astron. Astrophys.*, 604:A94, August 2017.
- [118] J. Erler, K. Basu, J. Chluba, and F. Bertoldi. Planck's view on the spectrum of the Sunyaev-Zeldovich effect. *Mon. Not. Roy. Astron. Soc.*, 476:3360–3381, May 2018.
- [119] Mathieu Remazeilles and Jens Chluba. Mapping the relativistic electron gas temperature across the sky. *Mon. Not. Roy. Astron. Soc.*, 494(4):5734–5750, June 2020.
- [120] Darren J. Croton, Volker Springel, Simon D. M. White, G. De Lucia, et al. The many lives of active galactic nuclei: cooling flows, black holes and the luminosities and colours of galaxies. *Mon. Not. Roy. Astron. Soc.*, 365(1):11–28, January 2006.
- [121] Rachel S. Somerville, Gergö Popping, and Scott C. Trager. Star formation in semi-analytic galaxy formation models with multiphase gas. *Mon. Not. Roy. Astron. Soc.*, 453(4):4337–4367, November 2015.

- [122] Volker Springel, Tiziana Di Matteo, and Lars Hernquist. Modelling feedback from stars and black holes in galaxy mergers. *Mon. Not. Roy. Astron. Soc.*, 361(3):776–794, August 2005.
- [123] Joop Schaye, Robert A. Crain, Richard G. Bower, Michelle Furlong, et al. The EAGLE project: simulating the evolution and assembly of galaxies and their environments. *Mon. Not. Roy. Astron. Soc.*, 446(1):521–554, January 2015.
- [124] Daniel Anglés-Alcázar, Romeel Davé, Claude-André Faucher-Giguère, Feryal Özel, et al. Gravitational torque-driven black hole growth and feedback in cosmological simulations. *Mon. Not. Roy. Astron. Soc.*, 464(3):2840–2853, January 2017.
- [125] S. Kaviraj, C. Laigle, T. Kimm, J. E. G. Devriendt, et al. The Horizon-AGN simulation: evolution of galaxy properties over cosmic time. *Mon. Not. Roy. Astron. Soc.*, 467(4):4739–4752, June 2017.
- [126] Philip F. Hopkins, Andrew Wetzel, Dušan Kereš, Claude-André Faucher-Giguère, et al. FIRE-2 simulations: physics versus numerics in galaxy formation. *Mon. Not. Roy. Astron. Soc.*, 480(1):800–863, October 2018.
- [127] Mark Vogelsberger, Federico Marinacci, Paul Torrey, and Ewald Puchwein. Cosmological simulations of galaxy formation. *Nature Reviews Physics*, 2(1):42–66, January 2020.
- [128] P. J. E. Peebles. The Black-Body Radiation Content of the Universe and the Formation of Galaxies. *Astrophys. J.*, 142:1317, November 1965.
- [129] William H. Press and Paul Schechter. Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation. *Astrophys. J.*, 187:425–438, February 1974.
- [130] Asantha Cooray and Ravi Sheth. Halo models of large scale structure. *Phys. Rep.*, 372(1):1–129, December 2002.
- [131] Digvijay Wadekar, Leander Thiele, Francisco Villaescusa-Navarro, J. Colin Hill, et al. Insights into baryonic feedback from the SZ flux–mass relation: Reducing deviations from self-similarity. *in prep.*, 2022.
- [132] Rainer Weinberger, Volker Springel, and Rüdiger Pakmor. The AREPO Public Code Release. *Astrophys. J. Suppl.*, 248(2):32, June 2020.
- [133] Volker Springel, Rüdiger Pakmor, Annalisa Pillepich, Rainer Weinberger, et al. First results from the IllustrisTNG simulations: matter and galaxy clustering. *Mon. Not. Roy. Astron. Soc.*, 475(1):676–698, March 2018.
- [134] Jill P. Naiman, Annalisa Pillepich, Volker Springel, Enrico Ramirez-Ruiz, et al. First results from the IllustrisTNG simulations: a tale of two elements - chemical evolution of magnesium and europium. *Mon. Not. Roy. Astron. Soc.*, 477(1):1206–1224, June 2018.

- [135] Dylan Nelson, Annalisa Pillepich, Volker Springel, Rainer Weinberger, et al. First results from the IllustrisTNG simulations: the galaxy colour bimodality. *Mon. Not. Roy. Astron. Soc.*, 475(1):624–647, March 2018.
- [136] Federico Marinacci, Mark Vogelsberger, Rüdiger Pakmor, Paul Torrey, et al. First results from the IllustrisTNG simulations: radio haloes and magnetic fields. *Mon. Not. Roy. Astron. Soc.*, 480(4):5113–5139, November 2018.
- [137] Annalisa Pillepich, Dylan Nelson, Lars Hernquist, Volker Springel, et al. First results from the IllustrisTNG simulations: the stellar mass content of groups and clusters of galaxies. *Mon. Not. Roy. Astron. Soc.*, 475(1):648–675, March 2018.
- [138] Dylan Nelson, Volker Springel, Annalisa Pillepich, Vicente Rodriguez-Gomez, et al. The IllustrisTNG simulations: public data release. *Computational Astrophysics and Cosmology*, 6(1):2, May 2019.
- [139] Philip F. Hopkins. A new class of accurate, mesh-free hydrodynamic simulation methods. *Mon. Not. Roy. Astron. Soc.*, 450(1):53–110, June 2015.
- [140] Romeel Davé, Daniel Anglés-Alcázar, Desika Narayanan, Qi Li, et al. SIMBA: Cosmological simulations with black hole growth and feedback. *Mon. Not. Roy. Astron. Soc.*, 486(2):2827–2849, June 2019.
- [141] A. Kogut and D. J. Fixsen. Calibration method and uncertainty for the primordial inflation explorer (PIXIE). *J. Cosmol. Astropart. Phys.*, 2020(5):041, May 2020.
- [142] Planck Collaboration et al. Planck 2015 results. X. Diffuse component separation: Foreground maps. *Astron. Astrophys.*, 594:A10, September 2016.
- [143] Aditya Rotti and Jens Chluba. Combining ILC and moment expansion techniques for extracting average-sky signals and CMB anisotropies. *Mon. Not. Roy. Astron. Soc.*, 500(1):976–985, January 2021.
- [144] Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, et al. Optuna: A Next-generation Hyperparameter Optimization Framework. *arXiv e-prints*, page arXiv:1907.10902, July 2019.
- [145] Jeremy L. Tinker, Brant E. Robertson, Andrey V. Kravtsov, Anatoly Klypin, et al. The Large-scale Bias of Dark Matter Halos: Numerical Calibration and Model Tests. *Astrophys. J.*, 724(2):878–886, December 2010.
- [146] M. Davis, G. Efstathiou, C. S. Frenk, and S. D. M. White. The evolution of large-scale structure in a universe dominated by cold dark matter. *Astrophys. J.*, 292:371–394, May 1985.
- [147] N. Battaglia, J. R. Bond, C. Pfrommer, and J. L. Sievers. On the Cluster Physics of Sunyaev-Zel'dovich and X-Ray Surveys. II. Deconstructing the Thermal SZ Power Spectrum. *Astrophys. J.*, 758(2):75, October 2012.

- [148] M. Arnaud, E. Pointecouteau, and G. W. Pratt. The structural and scaling properties of nearby galaxy clusters. II. The M-T relation. *Astron. Astrophys.*, 441(3):893–903, October 2005.
- [149] Greg L. Bryan and Michael L. Norman. Statistical Properties of X-Ray Clusters: Analytic and Numerical Comparisons. *Astrophys. J.*, 495(1):80–99, March 1998.
- [150] Josh Borrow, Daniel Anglés-Alcázar, and Romeel Davé. Cosmological baryon transfer in the SIMBA simulations. *Mon. Not. Roy. Astron. Soc.*, 491(4):6102–6119, February 2020.
- [151] Daniel Anglés-Alcázar, Claude-André Faucher-Giguère, Eliot Quataert, Philip F. Hopkins, et al. Black holes on FIRE: stellar feedback limits early feeding of galactic nuclei. *Mon. Not. Roy. Astron. Soc.*, 472(1):L109–L114, November 2017.
- [152] C. M. Booth and Joop Schaye. The interaction between feedback from active galactic nuclei and supernovae. *Scientific Reports*, page 1738, May 2013.
- [153] A. Amon, D. Gruen, M. A. Troxel, N. MacCrann, et al. Dark Energy Survey Year 3 results: Cosmology from cosmic shear and robustness to data calibration. *Phys. Rev. D*, 105(2):023514, January 2022.
- [154] Bhuvnesh Jain and Uroš Seljak. Cosmological Model Predictions for Weak Lensing: Linear and Nonlinear Regimes. *Astrophys. J.*, 484(2):560–573, July 1997.
- [155] Emily Moser, Nicholas Battaglia, Daisuke Nagai, Erwin Lau, et al. The Circumgalactic Medium from the CAMELS Simulations: Forecasting Constraints on Feedback Processes from Future Sunyaev-Zeldovich Observations. *arXiv e-prints*, page arXiv:2201.02708, January 2022.
- [156] Leander Thiele, Francisco Villaescusa-Navarro, David N. Spergel, Dylan Nelson, et al. Teaching Neural Networks to Generate Fast Sunyaev-Zel'dovich Maps. *Astrophys. J.*, 902(2):129, October 2020.
- [157] Roberto Scaramella, Renyue Cen, and Jeremiah P. Ostriker. A Hydrodynamic Approach to Cosmology: Nonlinear Effects on Cosmic Backgrounds in the Cold Dark Matter Model. *Astrophys. J.*, 416:399, October 1993.
- [158] Antonio C. da Silva, Domingos Barbosa, Andrew R. Liddle, and Peter A. Thomas. Hydrodynamical simulations of the Sunyaev-Zel'dovich effect. *Mon. Not. Roy. Astron. Soc.*, 317(1):37–44, September 2000.
- [159] Antonio C. da Silva, Domingos Barbosa, Andrew R. Liddle, and Peter A. Thomas. Hydrodynamical simulations of the Sunyaev-Zel'dovich effect: the kinetic effect. *Mon. Not. Roy. Astron. Soc.*, 326(1):155–163, September 2001.
- [160] Volker Springel, Martin White, and Lars Hernquist. Hydrodynamic Simulations of the Sunyaev-Zeldovich Effect(s). *Astrophys. J.*, 549(2):681–687, March 2001.

- [161] Martin White, Lars Hernquist, and Volker Springel. Simulating the Sunyaev-Zeldovich Effect(s): Including Radiative Cooling and Energy Injection by Galactic Winds. *Astrophys. J.*, 579(1):16–22, November 2002.
- [162] Daisuke Nagai, Andrey V. Kravtsov, and Alexey Vikhlinin. Effects of Galaxy Formation on Thermodynamics of the Intracluster Medium. *Astrophys. J.*, 668(1):1–14, October 2007.
- [163] C. Pfrommer, T. A. Enßlin, V. Springel, M. Jubelgas, et al. Simulating cosmic rays in clusters of galaxies - I. Effects on the Sunyaev-Zel'dovich effect and the X-ray emission. *Mon. Not. Roy. Astron. Soc.*, 378(2):385–408, June 2007.
- [164] Debora Sijacki, Christoph Pfrommer, Volker Springel, and Torsten A. Enßlin. Simulations of cosmic-ray feedback by active galactic nuclei in galaxy clusters. *Mon. Not. Roy. Astron. Soc.*, 387(4):1403–1415, July 2008.
- [165] N. Battaglia, J. R. Bond, C. Pfrommer, J. L. Sievers, et al. Simulations of the Sunyaev-Zel'dovich Power Spectrum with Active Galactic Nucleus Feedback. *Astrophys. J.*, 725(1):91–99, December 2010.
- [166] E. J. Hallman, B. W. O’Shea, M. L. Norman, R. Wagner, et al. AMR Simulations of the Cosmological Light Cone: SZE Surveys of the Synthetic Universe. In H. Böhringer, G. W. Pratt, A. Finoguenov, and P. Schuecker, editors, *Heating versus Cooling in Galaxies and Clusters of Galaxies*, page 355, January 2007.
- [167] B. M. Schäfer, C. Pfrommer, M. Bartelmann, V. Springel, et al. Detecting Sunyaev-Zel'dovich clusters with Planck - I. Construction of all-sky thermal and kinetic SZ maps. *Mon. Not. Roy. Astron. Soc.*, 370(3):1309–1323, August 2006.
- [168] Scott T. Kay, Michael W. Peel, C. J. Short, Peter A. Thomas, et al. Sunyaev-Zel'dovich clusters in Millennium gas simulations. *Mon. Not. Roy. Astron. Soc.*, 422(3):1999–2023, May 2012.
- [169] Joop Schaye, Robert A. Crain, Richard G. Bower, Michelle Furlong, et al. The EAGLE project: simulating the evolution and assembly of galaxies and their environments. *Mon. Not. Roy. Astron. Soc.*, 446(1):521–554, January 2015.
- [170] K. Dolag, E. Komatsu, and R. Sunyaev. SZ effects in the Magneticum Pathfinder simulation: comparison with the Planck, SPT, and ACT results. *Mon. Not. Roy. Astron. Soc.*, 463(2):1797–1811, December 2016.
- [171] Alexander Spacek, Mark L. A. Richardson, Evan Scannapieco, Julien Devriendt, et al. Using Real and Simulated Measurements of the Thermal Sunyaev-Zel'dovich Effect to Constrain Models of AGN Feedback. *Astrophys. J.*, 865(2):109, October 2018.
- [172] Dylan Nelson, Volker Springel, Annalisa Pillepich, Vicente Rodriguez-Gomez, et al. The IllustrisTNG simulations: public data release. *Computational Astrophysics and Cosmology*, 6(1):2, May 2019.

- [173] Fred M. Persi, David N. Spergel, Renyue Cen, and J. P. Ostriker. Hot Gas in Superclusters and Microwave Background Distortions. *Astrophys. J.*, 442:1, March 1995.
- [174] Alexandre Refregier, Eiichiro Komatsu, David N. Spergel, and Ue-Li Pen. Power spectrum of the Sunyaev-Zel'dovich effect. *Phys. Rev. D*, 61(12):123001, June 2000.
- [175] Eiichiro Komatsu and Tetsu Kitayama. Sunyaev-Zeldovich Fluctuations from Spatial Correlations between Clusters of Galaxies. *Astrophys. J. Lett.*, 526(1):L1–L4, November 1999.
- [176] E. Komatsu and U. Seljak. The Sunyaev-Zel'dovich angular power spectrum as a probe of cosmological parameters. *Mon. Not. Roy. Astron. Soc.*, 336(4):1256–1270, November 2002.
- [177] E. Komatsu and U. Seljak. Universal gas density and temperature profile. *Mon. Not. Roy. Astron. Soc.*, 327(4):1353–1366, November 2001.
- [178] Jounghun Lee and Yasushi Suto. Modeling Intracluster Gas in Triaxial Dark Halos: An Analytic Approach. *Astrophys. J.*, 585(1):151–160, March 2003.
- [179] Jeremiah P. Ostriker, Paul Bode, and Arif Babul. A Simple and Accurate Model for Intracluster Gas. *Astrophys. J.*, 634(2):964–976, December 2005.
- [180] Eric J. Hallman, Jack O. Burns, Patrick M. Motl, and Michael L. Norman. The β -Model Problem: The Incompatibility of X-Ray and Sunyaev-Zeldovich Effect Model Fitting for Galaxy Clusters. *Astrophys. J.*, 665(2):911–920, August 2007.
- [181] Paul Bode, Jeremiah P. Ostriker, Jochen Weller, and Laurie Shaw. Accurate Realizations of the Ionized Gas in Galaxy Clusters: Calibrating Feedback. *Astrophys. J.*, 663(1):139–149, July 2007.
- [182] Paul Bode, Jeremiah P. Ostriker, and Alexey Vikhlinin. Exploring the Energetics of Intracluster Gas with a Simple and Accurate Model. *Astrophys. J.*, 700(2):989–999, August 2009.
- [183] Laurie D. Shaw, Daisuke Nagai, Suman Bhattacharya, and Erwin T. Lau. Impact of Cluster Physics on the Sunyaev-Zel'dovich Power Spectrum. *Astrophys. J.*, 725(2):1452–1465, December 2010.
- [184] George Efstathiou and Marina Migliaccio. A simple empirically motivated template for the thermal Sunyaev-Zel'dovich effect. *Mon. Not. Roy. Astron. Soc.*, 423(3):2492–2497, July 2012.
- [185] Pedro R. Capelo, Paolo S. Coppi, and Priyamvada Natarajan. The polytropic approximation and X-ray scaling relations: constraints on gas and dark matter profiles for galaxy groups and clusters. *Mon. Not. Roy. Astron. Soc.*, 422(1):686–703, May 2012.

- [186] Xun Shi and Eiichiro Komatsu. Analytical model for non-thermal pressure in galaxy clusters. *Mon. Not. Roy. Astron. Soc.*, 442(1):521–532, July 2014.
- [187] Hy Trac, Paul Bode, and Jeremiah P. Ostriker. Templates for the Sunyaev-Zel'dovich Angular Power Spectrum. *Astrophys. J.*, 727(2):94, February 2011.
- [188] James R. Allison, Angela C. Taylor, Michael E. Jones, Steve Rawlings, et al. A parametric physical model for the intracluster medium and its use in joint SZ/X-ray analyses of galaxy clusters. *Mon. Not. Roy. Astron. Soc.*, 410(1):341–358, January 2011.
- [189] Kaylea Nelson, Erwin T. Lau, and Daisuke Nagai. Hydrodynamic Simulation of Non-thermal Pressure Profiles of Galaxy Clusters. *Astrophys. J.*, 792(1):25, September 2014.
- [190] Amandine M. C. Le Brun, Ian G. McCarthy, Joop Schaye, and Trevor J. Ponman. The scatter and evolution of the global hot gas properties of simulated galaxy cluster populations. *Mon. Not. Roy. Astron. Soc.*, 466(4):4442–4469, April 2017.
- [191] N. Gupta, A. Saro, J. J. Mohr, K. Dolag, et al. SZE observables, pressure profiles and centre offsets in Magneticum simulation galaxy clusters. *Mon. Not. Roy. Astron. Soc.*, 469(3):3069–3087, August 2017.
- [192] S. Planelles, D. Fabjan, S. Borgani, G. Murante, et al. Pressure of the hot gas in simulations of galaxy clusters. *Mon. Not. Roy. Astron. Soc.*, 467(4):3827–3847, June 2017.
- [193] A. J. Mead, T. Tröster, C. Heymans, L. Van Waerbeke, et al. A hydrodynamical halo model for weak-lensing cross correlations. *arXiv e-prints*, page arXiv:2005.00009, April 2020.
- [194] Niayesh Afshordi, Yen-Ting Lin, Daisuke Nagai, and Alastair J. R. Sanderson. Missing thermal energy of the intracluster medium. *Mon. Not. Roy. Astron. Soc.*, 378(1):293–300, June 2007.
- [195] F. Atrio-Barandela, A. Kashlinsky, D. Kocevski, and H. Ebeling. Measurement of the Electron-Pressure Profile of Galaxy Clusters in 3 Year Wilkinson Microwave Anisotropy Probe (WMAP) Data. *Astrophys. J. Lett.*, 675(2):L57, March 2008.
- [196] M. Arnaud, G. W. Pratt, R. Piffaretti, H. Böhringer, et al. The universal galaxy cluster pressure profile from a representative sample of nearby systems (REXCESS) and the Y_{SZ} - M_{500} relation. *Astron. Astrophys.*, 517:A92, July 2010.
- [197] Anya Chaudhuri and Subhabrata Majumdar. Sunyaev-Zel'dovich Scaling Relations from a Simple Phenomenological Model for Galaxy Clusters. *Astrophys. J. Lett.*, 728(2):L41, February 2011.

- [198] Fabio Zandanel, Christoph Pfrommer, and Francisco Prada. A phenomenological model for the intracluster medium that matches X-ray and Sunyaev-Zel'dovich observations. *Mon. Not. Roy. Astron. Soc.*, 438(1):116–123, February 2014.
- [199] Amandine M. C. Le Brun, Ian G. McCarthy, Joop Schaye, and Trevor J. Ponman. Towards a realistic population of simulated galaxy groups and clusters. *Mon. Not. Roy. Astron. Soc.*, 441(2):1270–1290, June 2014.
- [200] M. E. Ramos-Ceja, K. Basu, F. Pacaud, and F. Bertoldi. Constraining the intracluster pressure profile from the thermal SZ power spectrum. *Astron. Astrophys.*, 583:A111, November 2015.
- [201] J. P. Ostriker and E. T. Vishniac. Generation of Microwave Background Fluctuations from Nonlinear Perturbations at the ERA of Galaxy Formation. *Astrophys. J. Lett.*, 306:L51, July 1986.
- [202] Ethan T. Vishniac. Reionization and Small-Scale Fluctuations in the Microwave Background. *Astrophys. J.*, 322:597, November 1987.
- [203] Andrew H. Jaffe and Marc Kamionkowski. Calculation of the Ostriker-Vishniac effect in cold dark matter models. *Phys. Rev. D*, 58(4):043001, August 1998.
- [204] P. Valageas, A. Balbi, and J. Silk. Secondary CMB anisotropies from the kinetic SZ effect. *Astron. Astrophys.*, 367:1–17, February 2001.
- [205] Chung-Pei Ma and J. N. Fry. Nonlinear Kinetic Sunyaev-Zeldovich Effect. *Phys. Rev. Lett.*, 88(21):211301, May 2002.
- [206] Pengjie Zhang, Ue-Li Pen, and Hy Trac. Precision era of the kinetic Sunyaev-Zel'dovich effect: simulations, analytical models and observations and the power to constrain reionization. *Mon. Not. Roy. Astron. Soc.*, 347(4):1224–1233, February 2004.
- [207] Hyunbae Park, Eiichiro Komatsu, Paul R. Shapiro, Jun Koda, et al. The Impact of Nonlinear Structure Formation on the Power Spectrum of Transverse Momentum Fluctuations and the Kinetic Sunyaev-Zel'dovich Effect. *Astrophys. J.*, 818(1):37, February 2016.
- [208] Laurie D. Shaw, Douglas H. Rudd, and Daisuke Nagai. Deconstructing the Kinetic SZ Power Spectrum. *Astrophys. J.*, 756(1):15, September 2012.
- [209] Marcelo A. Alvarez. The Kinetic Sunyaev-Zel'dovich Effect from Reionization: Simulated Full-sky Maps at Arcminute Resolution. *Astrophys. J.*, 824(2):118, June 2016.
- [210] Hyunbae Park, Marcelo A. Alvarez, and J. Richard Bond. The Impact of Baryonic Physics on the Kinetic Sunyaev-Zel'dovich Effect. *Astrophys. J.*, 853(2):121, February 2018.

- [211] Tilman Tröster, Cameron Ferguson, Joachim Harnois-Déraps, and Ian G. McCarthy. Painting with baryons: augmenting N-body simulations with gas using deep generative models. *Mon. Not. Roy. Astron. Soc.*, 487(1):L24–L29, July 2019.
- [212] Xiaoying Xu, Shirley Ho, Hy Trac, Jeff Schneider, et al. A First Look at Creating Mock Catalogs with Machine Learning Techniques. *Astrophys. J.*, 772(2):147, August 2013.
- [213] Xinyue Zhang, Yanfang Wang, Wei Zhang, Yueqiu Sun, et al. From Dark Matter to Galaxies with Convolutional Networks. *arXiv e-prints*, page arXiv:1902.05965, February 2019.
- [214] Shankar Agarwal, Romeel Davé, and Bruce A. Bassett. Painting galaxies into dark matter haloes using machine learning. *Mon. Not. Roy. Astron. Soc.*, 478(3):3410–3422, August 2018.
- [215] Yongseok Jo and Ji-hoon Kim. Machine-assisted semi-simulation model (MSSM): estimating galactic baryonic properties from their dark matter using a machine trained on hydrodynamic simulations. *Mon. Not. Roy. Astron. Soc.*, 489(3):3565–3581, November 2019.
- [216] Benjamin P. Moster, Thorsten Naab, Magnus Lindström, and Joseph A. O’Leary. GalaxyNet: Connecting galaxies and dark matter haloes with deep neural networks and reinforcement learning in large volumes. *arXiv e-prints*, page arXiv:2005.12276, May 2020.
- [217] Siyu He, Yin Li, Yu Feng, Shirley Ho, et al. Learning to predict the cosmological structure formation. *Proceedings of the National Academy of Science*, 116(28):13825–13832, July 2019.
- [218] Rainer Weinberger, Volker Springel, Lars Hernquist, Annalisa Pillepich, et al. Simulating galaxy formation with black hole driven thermal and kinetic feedback. *Mon. Not. Roy. Astron. Soc.*, 465(3):3291–3308, March 2017.
- [219] Annalisa Pillepich, Volker Springel, Dylan Nelson, Shy Genel, et al. Simulating galaxy formation with the IllustrisTNG model. *Mon. Not. Roy. Astron. Soc.*, 473(3):4077–4106, January 2018.
- [220] N. Battaglia. The tau of galaxy clusters. *J. Cosmol. Astropart. Phys.*, 2016(8):058, August 2016.
- [221] Nelson et al. in prep.
- [222] Nick Hand, Yu Feng, Florian Beutler, Yin Li, et al. nbodykit: An Open-source, Massively Parallel Toolkit for Large-scale Structure. *Astron. J.*, 156(4):160, October 2018.

- [223] Francisco Villaescusa-Navarro. Pylians: Python libraries for the analysis of numerical simulations, November 2018.
- [224] A. R. Duffy, J. Schaye, S. T. Kay, and C. Dalla Vecchia. Dark matter halo concentrations in the Wilkinson Microwave Anisotropy Probe year 5 cosmology. *Mon. Not. Roy. Astron. Soc.*, 390:L64–L68, October 2008.
- [225] J. F. Navarro, C. S. Frenk, and S. D. M. White. A Universal Density Profile from Hierarchical Clustering. *Astrophys. J.*, 490:493–508, December 1997.
- [226] J. Colin Hill and David N. Spergel. Detection of thermal SZ-CMB lensing cross-correlation in Planck nominal mission data. *J. Cosmol. Astropart. Phys.*, 2014(2):030, February 2014.
- [227] Ludovic Van Waerbeke, Gary Hinshaw, and Norman Murray. Detection of warm and diffuse baryons in large scale structure from the cross correlation of gravitational lensing and the thermal Sunyaev-Zeldovich effect. *Phys. Rev. D*, 89(2):023508, January 2014.
- [228] Alireza Hojjati, Tilman Tröster, Joachim Harnois-Déraps, Ian G. McCarthy, et al. Cross-correlating Planck tSZ with RCSLenS weak lensing: implications for cosmology and AGN feedback. *Mon. Not. Roy. Astron. Soc.*, 471(2):1565–1580, October 2017.
- [229] S. Nozawa, N. Itoh, Y. Suda, and Y. Ohhata. An improved formula for the relativistic corrections to the kinematical Sunyaev-Zeldovich effect for clusters of galaxies. *Nuovo Cimento B Serie*, 121(5):487–500, May 2006.
- [230] Mathieu Remazeilles and Jens Chluba. Mapping the relativistic electron gas temperature across the sky. *Mon. Not. Roy. Astron. Soc.*, 494(4):5734–5750, May 2020.
- [231] Elizabeth Lee, Jens Chluba, Scott T. Kay, and David J. Barnes. Relativistic SZ temperature scaling relations of groups and clusters derived from the BAHAMAS and MACSIS simulations. *Mon. Not. Roy. Astron. Soc.*, 493(3):3274–3292, April 2020.
- [232] Villaescusa-Navarro et al. in prep.
- [233] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, et al. Generative Adversarial Networks. *arXiv e-prints*, page arXiv:1406.2661, June 2014.
- [234] Naftali Tishby, Esther Levin, and Sara A. Solla. Consistent inference of probabilities in layered networks: Predictions and generalization. In Anon, editor, *IJCNN Int Jt Conf Neural Network*, pages 403–409. Publ by IEEE, December 1989. IJCNN International Joint Conference on Neural Networks ; Conference date: 18-06-1989 Through 22-06-1989.

- [235] Justin Alsing, Benjamin Wandelt, and Stephen Feeney. Massive optimal data compression and density estimation for scalable, likelihood-free inference in cosmology. *Mon. Not. Roy. Astron. Soc.*, 477:2874–2885, July 2018.
- [236] J. Colin Hill, Blake D. Sherwin, Kendrick M. Smith, Graeme E. Addison, et al. The Atacama Cosmology Telescope: A Measurement of the Thermal Sunyaev-Zel’dovich One-Point PDF. *arXiv e-prints*, page arXiv:1411.8004, November 2014.
- [237] Leander Thiele, J. Colin Hill, and Kendrick M. Smith. Accurate analytic model for the thermal Sunyaev-Zel’dovich one-point probability distribution function. *Phys. Rev. D*, 99(10):103511, May 2019.
- [238] Leander Thiele, Miles Cranmer, William Coulton, Shirley Ho, et al. Predicting the thermal Sunyaev-Zel’dovich field using modular and equivariant set-based neural networks. *Machine Learning: Science and Technology*, 3(3):035002, September 2022.
- [239] R. A. Sunyaev and Ia. B. Zeldovich. Microwave background radiation as a probe of the contemporary structure and history of the universe. *Annu. Rev. Astron. Astrophys.*, 18:537–560, January 1980.
- [240] M. Gatti, S. Pandey, E. Baxter, J. C. Hill, et al. Cross-correlation of DES Y3 lensing and ACT/Planck thermal Sunyaev Zel’dovich Effect I: Measurements, systematics tests, and feedback model constraints. *arXiv e-prints*, page arXiv:2108.01600, August 2021.
- [241] Tilman Tröster, Cameron Ferguson, Joachim Harnois-Déraps, and Ian G. McCarthy. Painting with baryons: augmenting N-body simulations with gas using deep generative models. *Mon. Not. Roy. Astron. Soc.*, 487(1):L24–L29, July 2019.
- [242] Jacky H. T. Yip, Xinyue Zhang, Yanfang Wang, Wei Zhang, et al. From Dark Matter to Galaxies with Convolutional Neural Networks. *arXiv e-prints*, page arXiv:1910.07813, October 2019.
- [243] Xinyue Zhang, Yanfang Wang, Wei Zhang, Yueqiu Sun, et al. From Dark Matter to Galaxies with Convolutional Networks. *arXiv e-prints*, page arXiv:1902.05965, February 2019.
- [244] Noah Kasmanoff, Francisco Villaescusa-Navarro, Jeremy Tinker, and Shirley Ho. dm2gal: Mapping Dark Matter to Galaxies with Neural Networks. *arXiv e-prints*, page arXiv:2012.00186, November 2020.
- [245] Digvijay Wadekar, Francisco Villaescusa-Navarro, Shirley Ho, and Laurence Perreault-Levasseur. HInet: Generating Neutral Hydrogen from Dark Matter with Neural Networks. *Astrophys. J.*, 916(1):42, July 2021.

- [246] Tibor Rothschild, Daisuke Nagai, Han Aung, Sheridan B. Green, et al. Emulating sunyaev-zeldovich images of galaxy clusters using auto-encoders, 2021.
- [247] Daisuke Nagai, Andrey V. Kravtsov, and Alexey Vikhlinin. Effects of Galaxy Formation on Thermodynamics of the Intracluster Medium. *Astrophys. J.*, 668(1):1–14, October 2007.
- [248] Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, et al. Deep Sets. *arXiv e-prints*, page arXiv:1703.06114, March 2017.
- [249] Michael M. Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, et al. Geometric Deep Learning: Going beyond Euclidean data. *IEEE Signal Processing Magazine*, 34(4):18–42, July 2017.
- [250] Peter Battaglia, Razvan Pascanu, Matthew Lai, Danilo Jimenez Rezende, et al. Interaction networks for learning about objects, relations and physics. In *Advances in Neural Information Processing Systems*, pages 4502–4510, 2016.
- [251] Peter W. Battaglia, Jessica B. Hamrick, Victor Bapst, Alvaro Sanchez-Gonzalez, et al. Relational inductive biases, deep learning, and graph networks. *arXiv e-prints*, page arXiv:1806.01261, June 2018.
- [252] Michael M. Bronstein, Joan Bruna, Taco Cohen, and Petar Veličković. Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges. *arXiv:2104.13478 [cs, stat]*, May 2021.
- [253] Ademola Oladosu, Tony Xu, Philip Ekfeldt, Brian A. Kelly, et al. Meta-Learning for One-Class Classification with Few Examples using Order-Equivariant Network. *arXiv:2007.04459 [astro-ph, stat]*, May 2021.
- [254] Miles Cranmer, Alvaro Sanchez-Gonzalez, Peter Battaglia, Rui Xu, et al. Discovering Symbolic Models from Deep Learning with Inductive Biases. *NeurIPS*, June 2020.
- [255] Miles Cranmer, Christina Kreisch, Alice Pisani, Francisco Villaescusa-Navarro, et al. Histogram Pooling Operators: An Interpretable Alternative for DeepSets. In *Deep Learning for Simulation, ICLR 2021 Workshop*, page 10, 2021.
- [256] Miles Cranmer, Peter Melchior, and Brian Nord. Unsupervised Resource Allocation with Graph Neural Networks. *arXiv:2106.09761 [astro-ph]*, pages 1–13, June 2021.
- [257] Soledad Villar, David W. Hogg, Kate Storey-Fisher, Weichi Yao, et al. Scalars are universal: Gauge-equivariant machine learning, structured like classical physics. *arXiv e-prints*, page arXiv:2106.06610, June 2021.
- [258] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. *arXiv e-prints*, page arXiv:1603.08155, 2016.

- [259] Patrick Esser, Ekaterina Sutter, and Björn Ommer. A variational u-net for conditional appearance and shape generation. *arXiv e-prints*, page arXiv:1804.04694, 2018.
- [260] François Lanusse, Rachel Mandelbaum, Siamak Ravanbakhsh, Chun-Liang Li, et al. Deep generative models for galaxy image simulations. *Mon. Not. Roy. Astron. Soc.*, 504(4):5543–5555, July 2021.
- [261] Benjamin Horowitz, Max Dornfest, Zarija Lukić, and Peter Harrington. HyPhy: Deep Generative Conditional Posterior Mapping of Hydrodynamical Physics. *arXiv e-prints*, page arXiv:2106.12675, June 2021.
- [262] Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E. Hinton. Layer normalization, 2016.
- [263] Geoffrey E. Hinton, Nitish Srivastava, Alex Krizhevsky, Ilya Sutskever, et al. Improving neural networks by preventing co-adaptation of feature detectors. *CoRR*, abs/1207.0580, 2012.
- [264] Peter Behroozi, Risa Wechsler, and Hao-Yi Wu. Rockstar: Phase-space halo finder, October 2012.
- [265] Peter S. Behroozi, Risa H. Wechsler, and Hao-Yi Wu. The ROCKSTAR Phase-space Temporal Halo Finder and the Velocity Offsets of Cluster Cores. *Astrophys. J.*, 762(2):109, January 2013.
- [266] Francisco Villaescusa-Navarro, Shy Genel, Daniel Angles-Alcazar, Leander Thiele, et al. The CAMELS Multifield Dataset: Learning the Universe’s Fundamental Parameters with Artificial Intelligence. *arXiv e-prints*, page arXiv:2109.10915, September 2021.
- [267] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In Yoshua Bengio and Yann LeCun, editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.
- [268] Leslie N. Smith and Nicholay Topin. Super-convergence: Very fast training of residual networks using large learning rates. *CoRR*, abs/1708.07120, 2017.
- [269] Loïc Lannelongue, Jason Grealey, and Michael Inouye. Green algorithms: Quantifying the carbon footprint of computation. *Advanced Science*, 8(12):2100707, 2021.
- [270] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, et al. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d’Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019.

- [271] Charles R. Harris, K. Jarrod Millman, St  fan J. van der Walt, Ralf Gommers, et al. Array programming with NumPy. *Nature*, 585(7825):357–362, September 2020.
- [272] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020.
- [273] Thomas Kluyver, Benjamin Ragan-Kelley, Fernando P  rez, Brian Granger, et al. Jupyter Notebooks – a publishing format for reproducible computational workflows. In F. Loizides and B. Schmidt, editors, *Positioning and Power in Academic Publishing: Players, Agents and Agendas*, pages 87 – 90. IOS Press, 2016.
- [274] Fabian Pedregosa, Ga  l Varoquaux, Alexandre Gramfort, Vincent Michel, et al. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12(85):2825–2830, 2011.
- [275] J. D. Hunter. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, 9(3):90–95, 2007.
- [276] Matthias Fey and Jan E. Lenssen. Fast graph representation learning with PyTorch Geometric. In *ICLR Workshop on Representation Learning on Graphs and Manifolds*, 2019.
- [277] Astropy Collaboration, A. M. Price-Whelan, B. M. Sip  cz, H. M. G  nther, et al. The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package. *Astron. J.*, 156(3):123, September 2018.
- [278] Leander Thiele, Gabriela A. Marques, Jia Liu, and Masato Shirasaki. Cosmological constraints from the Subaru Hyper Suprime-Cam year 1 shear catalogue lensing convergence probability distribution function. *Phys. Rev. D*, 108(12):123526, December 2023.
- [279] Simone Aiola, Erminia Calabrese, Lo  c Maurin, Sigurd Naess, et al. The Atacama Cosmology Telescope: DR4 maps and cosmological parameters. *J. Cosmol. Astropart. Phys.*, 2020(12):047, December 2020.
- [280] Marika Asgari, Chieh-An Lin, Benjamin Joachimi, Benjamin Giblin, et al. KiDS-1000 cosmology: Cosmic shear constraints and comparison between two point statistics. *Astron. Astrophys.*, 645:A104, January 2021.
- [281] L. F. Secco, S. Samuroff, E. Krause, B. Jain, et al. Dark Energy Survey Year 3 results: Cosmology from cosmic shear and robustness to modeling uncertainty. *Phys. Rev. D*, 105(2):023515, January 2022.
- [282] Roohi Dalal, Xiangchong Li, Andrina Nicola, Joe Zuntz, et al. Hyper Suprime-Cam Year 3 Results: Cosmology from Cosmic Shear Power Spectra. April 2023.

- [283] Xiangchong Li, Tianqing Zhang, Sinao Sugiyama, Roohi Dalal, et al. Hyper Suprime-Cam Year 3 Results: Cosmology from Cosmic Shear Two-point Correlation Functions. April 2023.
- [284] Bhuvnesh Jain and Ludovic Van Waerbeke. Statistics of Dark Matter Halos from Gravitational Lensing. *Astrophys. J. Lett.*, 530(1):L1–L4, February 2000.
- [285] Laura Marian, Robert E. Smith, and Gary M. Bernstein. The Cosmology Dependence of Weak Lensing Cluster Counts. *Astrophys. J. Lett.*, 698(1):L33–L36, June 2009.
- [286] J. P. Dietrich and J. Hartlap. Cosmology with the shear-peak statistics. *Mon. Not. Roy. Astron. Soc.*, 402(2):1049–1058, February 2010.
- [287] Jan M. Kratochvil, Zoltán Haiman, and Morgan May. Probing cosmology with weak lensing peak counts. *Phys. Rev. D*, 81(4):043519, February 2010.
- [288] Jia Liu, Andrea Petri, Zoltán Haiman, Lam Hui, et al. Cosmology constraints from the weak lensing peak counts and the power spectrum in CFHTLenS data. *Phys. Rev. D*, 91(6):063507, March 2015.
- [289] Austin Peel, Chieh-An Lin, François Lanusse, Adrienne Leonard, et al. Cosmological constraints with weak-lensing peak counts and second-order statistics in a large-field survey. *Astron. Astrophys.*, 599:A79, March 2017.
- [290] Nicolas Martinet, Peter Schneider, Hendrik Hildebrandt, HuanYuan Shan, et al. KiDS-450: cosmological constraints from weak-lensing peak statistics - II: Inference from shear peaks using N-body simulations. *Mon. Not. Roy. Astron. Soc.*, 474(1):712–730, February 2018.
- [291] Joachim Harnois-Déraps, Nicolas Martinet, Tiago Castro, Klaus Dolag, et al. Cosmic shear cosmology beyond two-point statistics: a combined peak count and correlation function analysis of DES-Y1. *Mon. Not. Roy. Astron. Soc.*, 506(2):1623–1650, September 2021.
- [292] Alina Sabyr, Zoltán Haiman, José Manuel Zorrilla Matilla, and Tianhuan Lu. Cosmological constraints from weak lensing peaks: Can halo models accurately predict peak counts? *Phys. Rev. D*, 105(2):023505, January 2022.
- [293] Christopher T. Davies, Marius Cautun, Benjamin Giblin, Baojiu Li, et al. Constraining cosmology with weak lensing voids. *Mon. Not. Roy. Astron. Soc.*, 507(2):2267–2282, October 2021.
- [294] Dipak Munshi, Ludovic van Waerbeke, Joseph Smidt, and Peter Coles. From weak lensing to non-Gaussianity via Minkowski functionals. *Mon. Not. Roy. Astron. Soc.*, 419(1):536–555, January 2012.

- [295] Jan M. Kratochvil, Eugene A. Lim, Sheng Wang, Zoltán Haiman, et al. Probing cosmology with weak lensing Minkowski functionals. *Phys. Rev. D*, 85(10):103513, May 2012.
- [296] Andrea Petri, Zoltán Haiman, Lam Hui, Morgan May, et al. Cosmology with Minkowski functionals and moments of the weak lensing convergence field. *Phys. Rev. D*, 88(12):123002, December 2013.
- [297] Andrea Petri, Jia Liu, Zoltán Haiman, Morgan May, et al. Emulating the CFHTLenS weak lensing data: Cosmological constraints from moments and Minkowski functionals. *Phys. Rev. D*, 91(10):103511, May 2015.
- [298] Martina Vicinanza, Vincenzo F. Cardone, Roberto Maoli, Roberto Scaramella, et al. Minkowski functionals of convergence maps and the lensing figure of merit. *Phys. Rev. D*, 99(4):043534, February 2019.
- [299] Gabriela A Marques, Jia Liu, José Manuel Zorrilla Matilla, Zoltán Haiman, et al. Constraining neutrino mass with weak lensing minkowski functionals. *Journal of Cosmology and Astroparticle Physics*, 2019(06):019, 2019.
- [300] Carolina Parroni, Vincenzo F. Cardone, Roberto Maoli, and Roberto Scaramella. Going deep with Minkowski functionals of convergence maps. *Astron. Astrophys.*, 633:A71, January 2020.
- [301] Job Feldbrugge, Matti van Engelen, Rien van de Weygaert, Pratyush Pranav, et al. Stochastic homology of Gaussian vs. non-Gaussian random fields: graphs towards Betti numbers and persistence diagrams. *J. Cosmol. Astropart. Phys.*, 2019(9):052, September 2019.
- [302] Carolina Parroni, Édouard Tollet, Vincenzo F. Cardone, Roberto Maoli, et al. Higher-order statistics of shear field via a machine learning approach. *Astron. Astrophys.*, 645:A123, January 2021.
- [303] Sven Heydenreich, Benjamin Brück, and Joachim Harnois-Déraps. Persistent homology in cosmic shear: Constraining parameters with topological data analysis. *Astron. Astrophys.*, 648:A74, April 2021.
- [304] Sven Heydenreich, Benjamin Brück, Pierre Burger, Joachim Harnois-Déraps, et al. Persistent homology in cosmic shear. II. A tomographic analysis of DES-Y1. *Astron. Astrophys.*, 667:A125, November 2022.
- [305] S. Pires, J. L. Starck, A. Amara, A. Réfrégier, et al. Cosmological model discrimination with weak lensing. *Astron. Astrophys.*, 505(3):969–979, October 2009.
- [306] Sihao Cheng, Yuan-Sen Ting, Brice Ménard, and Joan Bruna. A new approach to observational cosmology using the scattering transform. *Mon. Not. Roy. Astron. Soc.*, 499(4):5902–5914, December 2020.

- [307] Siyao Cheng and Brice Ménard. How to quantify fields or textures? A guide to the scattering transform. *arXiv e-prints*, page arXiv:2112.01288, November 2021.
- [308] Masahiro Takada and Bhuvnesh Jain. The kurtosis of the cosmic shear field. *Mon. Not. Roy. Astron. Soc.*, 337(3):875–894, December 2002.
- [309] M. Kilbinger and P. Schneider. Cosmological parameters from combined second- and third-order aperture mass statistics of cosmic shear. *Astron. Astrophys.*, 442(1):69–83, October 2005.
- [310] L. Van Waerbeke, J. Benjamin, T. Erben, C. Heymans, et al. CFHTLenS: mapping the large-scale structure with gravitational lensing. *Mon. Not. Roy. Astron. Soc.*, 433(4):3373–3388, August 2013.
- [311] Lucas Porth and Robert E. Smith. Fast estimation of aperture-mass statistics - II. Detectability of higher order statistics in current and future surveys. *Mon. Not. Roy. Astron. Soc.*, 508(3):3474–3494, December 2021.
- [312] M. Gatti, B. Jain, C. Chang, M. Raveri, et al. Dark Energy Survey Year 3 results: Cosmology with moments of weak lensing mass maps. *Phys. Rev. D*, 106(8):083509, October 2022.
- [313] Asantha Cooray and Wayne Hu. Weak Gravitational Lensing Bispectrum. *Astrophys. J.*, 548(1):7–18, February 2001.
- [314] Masahiro Takada and Bhuvnesh Jain. Three-point correlations in weak lensing surveys: model predictions and applications. *Mon. Not. Roy. Astron. Soc.*, 344(3):857–886, September 2003.
- [315] Masahiro Takada and Bhuvnesh Jain. Cosmological parameters from lensing power spectrum and bispectrum tomography. *Mon. Not. Roy. Astron. Soc.*, 348(3):897–915, March 2004.
- [316] Scott Dodelson and Pengjie Zhang. Weak lensing bispectrum. *Phys. Rev. D*, 72(8):083001, October 2005.
- [317] Dragan Huterer, Masahiro Takada, Gary Bernstein, and Bhuvnesh Jain. Systematic errors in future weak-lensing surveys: requirements and prospects for self-calibration. *Mon. Not. Roy. Astron. Soc.*, 366(1):101–114, February 2006.
- [318] Sanaz Vafaei, Tingting Lu, Ludovic van Waerbeke, Elisabetta Semboloni, et al. Breaking the degeneracy: Optimal use of three-point weak lensing statistics. *Astroparticle Physics*, 32(6):340–351, January 2010.
- [319] D. Gruen, O. Friedrich, E. Krause, J. DeRose, et al. Density split statistics: Cosmological constraints from counts and lensing in cells in DES Y1 and SDSS data. *Phys. Rev. D*, 98(2):023507, July 2018.

- [320] Janis Fluri, Tomasz Kacprzak, Aurelien Lucchi, Alexandre Refregier, et al. Cosmological constraints with deep learning from KiDS-450 weak lensing maps. *Phys. Rev. D*, 100(6):063514, September 2019.
- [321] Janis Fluri, Tomasz Kacprzak, Aurelien Lucchi, Aurel Schneider, et al. Full wCDM analysis of KiDS-1000 weak lensing maps using deep learning. *Phys. Rev. D*, 105(8):083518, April 2022.
- [322] Tianhuan Lu, Zoltán Haiman, and Xiangchong Li. Cosmological constraints from HSC survey first-year data using deep learning. *Mon. Not. Roy. Astron. Soc.*, March 2023.
- [323] Euclid Collaboration. Euclid Preparation XXIX: Forecasts for 10 different higher-order weak lensing statistics. *arXiv e-prints*, page arXiv:2301.12890, January 2023.
- [324] Natalia Porqueres, Alan Heavens, Daniel Mortlock, and Guilhem Lavaux. Lifting weak lensing degeneracies with a field-based likelihood. *Mon. Not. Roy. Astron. Soc.*, 509(3):3194–3202, January 2022.
- [325] Natalia Porqueres, Alan Heavens, Daniel Mortlock, Guilhem Lavaux, et al. Field-level inference of cosmic shear with intrinsic alignments and baryons. *arXiv e-prints*, page arXiv:2304.04785, April 2023.
- [326] Adrian Bayer, Uros Seljak, and Chirag Modi. Field-Level Inference with Microcanonical Langevin Monte Carlo. In *Machine Learning for Astrophysics. Workshop at the Fortieth International Conference on Machine Learning (ICML 2023)*, page 3, July 2023.
- [327] Biwei Dai and Uros Seljak. Multiscale Flow for Robust and Optimal Cosmological Analysis. In *Machine Learning for Astrophysics. Workshop at the Fortieth International Conference on Machine Learning (ICML 2023)*, page 10, July 2023.
- [328] D. Zürcher, J. Fluri, V. Ajani, S. Fischbacher, et al. Towards a full wCDM map-based analysis for weak lensing surveys. *Mon. Not. Roy. Astron. Soc.*, 525(1):761–784, October 2023.
- [329] Željko Ivezić, Steven M. Kahn, J. Anthony Tyson, Bob Abel, et al. LSST: From Science Drivers to Reference Design and Anticipated Data Products. *Astrophys. J.*, 873(2):111, March 2019.
- [330] R. Laureijs, J. Amiaux, S. Arduini, J. L. Auguères, et al. Euclid Definition Study Report. *arXiv e-prints*, page arXiv:1110.3193, October 2011.
- [331] D. Spergel, N. Gehrels, C. Baltay, D. Bennett, et al. Wide-Field InfrarRed Survey Telescope-Astrophysics Focused Telescope Assets WFIRST-AFTA 2015 Report. *arXiv e-prints*, page arXiv:1503.03757, March 2015.

- [332] P. Valageas. Weak gravitational lensing effects on the determination of Omega_mega_m and Omega_mega_Lambda from SNeIa. *Astron. Astrophys.*, 354:767–786, February 2000.
- [333] P. Valageas. Statistical properties of the convergence due to weak gravitational lensing by non-linear structures. *Astron. Astrophys.*, 356:771–787, April 2000.
- [334] Dipak Munshi and Bhuvnesh Jain. The statistics of weak lensing at small angular scales: probability distribution function. *Mon. Not. Roy. Astron. Soc.*, 318(1):109–123, October 2000.
- [335] F. Bernardeau and P. Valageas. Construction of the one-point PDF of the local aperture mass in weak lensing maps. *Astron. Astrophys.*, 364:1–16, December 2000.
- [336] Hiroaki Aihara, Nobuo Arimoto, Robert Armstrong, Stéphane Arnouts, et al. The Hyper Suprime-Cam SSP Survey: Overview and survey design. *Publ. Astron. Soc. Jpn.*, 70:S4, January 2018.
- [337] Hiroaki Aihara, Robert Armstrong, Steven Bickerton, James Bosch, et al. First data release of the Hyper Suprime-Cam Subaru Strategic Program. *Publ. Astron. Soc. Jpn.*, 70:S8, January 2018.
- [338] Jia Liu, J. Colin Hill, Blake D. Sherwin, Andrea Petri, et al. CMB lensing beyond the power spectrum: Cosmological constraints from the one-point probability distribution function and peak counts. *Phys. Rev. D*, 94(10):103501, November 2016.
- [339] Kenneth Patton, Jonathan Blazek, Klaus Honscheid, Eric Huff, et al. Cosmological constraints from the convergence 1-point probability distribution. *Mon. Not. Roy. Astron. Soc.*, 472(1):439–446, November 2017.
- [340] Jia Liu and Mathew S. Madhavacheril. Constraining neutrino mass with the tomographic weak lensing one-point probability distribution function and power spectrum. *Phys. Rev. D*, 99(8):083508, April 2019.
- [341] Aoife Boyle, Cora Uhlemann, Oliver Friedrich, Alexandre Barthelemy, et al. Nuw CDM cosmology from the weak-lensing convergence PDF. *Mon. Not. Roy. Astron. Soc.*, 505(2):2886–2902, August 2021.
- [342] Benjamin Giblin, Yan-Chuan Cai, and Joachim Harnois-Déraps. Enhancing cosmic shear with the multiscale lensing probability density function. *Mon. Not. Roy. Astron. Soc.*, 520(2):1721–1737, April 2023.
- [343] Paulo Reimberg and Francis Bernardeau. Large deviation principle at work: Computation of the statistical properties of the exact one-point aperture mass. *Phys. Rev. D*, 97(2):023524, January 2018.

- [344] A. Barthelemy, S. Codis, C. Uhlemann, F. Bernardeau, et al. A nulling strategy for modelling lensing convergence in cones with large deviation theory. *Mon. Not. Roy. Astron. Soc.*, 492(3):3420–3439, March 2020.
- [345] Alexandre Barthelemy, Sandrine Codis, and Francis Bernardeau. Post-Born corrections to the one-point statistics of (CMB) lensing convergence obtained via large deviation theory. *Mon. Not. Roy. Astron. Soc.*, 494(3):3368–3382, May 2020.
- [346] Kimmo Kainulainen and Valerio Marra. Weak lensing observables in the halo model. *Phys. Rev. D*, 84(6):063004, September 2011.
- [347] Kimmo Kainulainen and Valerio Marra. Accurate modeling of weak lensing with the stochastic gravitational lensing method. *Phys. Rev. D*, 83(2):023009, January 2011.
- [348] Gabriela A. Marques, Jia Liu, Masato Shirasaki, Leander Thiele, et al. Cosmology from weak lensing peaks and minima with Subaru Hyper Suprime-Cam survey first-year data. *arXiv e-prints*, page arXiv:2308.10866, August 2023.
- [349] Rachel Mandelbaum, Hironao Miyatake, Takashi Hamana, Masamune Oguri, et al. The first-year shear catalog of the Subaru Hyper Suprime-Cam Subaru Strategic Program Survey. *Publ. Astron. Soc. Jpn.*, 70:S25, January 2018.
- [350] Masayuki Tanaka, Jean Coupon, Bau-Ching Hsieh, Sogo Mineo, et al. Photometric redshifts for Hyper Suprime-Cam Subaru Strategic Program Data Release 1. *Publ. Astron. Soc. Jpn.*, 70:S9, January 2018.
- [351] Nick Kaiser and Gordon Squires. Mapping the Dark Matter with Weak Gravitational Lensing. *Astrophys. J.*, 404:441, February 1993.
- [352] James Sunseri, Zack Li, and Jia Liu. Effects of baryonic feedback on the cosmic web. *Phys. Rev. D*, 107(2):023514, January 2023.
- [353] Eric Hivon, Krzysztof M Górski, C Barth Netterfield, Brendan P Crill, et al. Master of the cosmic microwave background anisotropy power spectrum: a fast method for statistical analysis of large and complex cosmic microwave background data sets. *The Astrophysical Journal*, 567(1):2, 2002.
- [354] David Alonso Monge, Javier Sanchez, and Anze Slosar. A unified pseudo-cl framework. *Monthly Notices of the Royal Astronomical Society*, 484(3):4127–4151, 2019.
- [355] Masamune Oguri, Satoshi Miyazaki, Chiaki Hikage, Rachel Mandelbaum, et al. Two- and three-dimensional wide-field weak lensing mass maps from the Hyper Suprime-Cam Subaru Strategic Program S16A data. *Publ. Astron. Soc. Jpn.*, 70:S26, January 2018.

- [356] Chiaki Hikage, Masamune Oguri, Takashi Hamana, Surhud More, et al. Cosmology from cosmic shear power spectra with Subaru Hyper Suprime-Cam first-year data. *Publ. Astron. Soc. Jpn.*, 71(2):43, April 2019.
- [357] Takashi Hamana, Masato Shirasaki, Satoshi Miyazaki, Chiaki Hikage, et al. Cosmological constraints from cosmic shear two-point correlation functions with HSC survey first-year data. *Publ. Astron. Soc. Jpn.*, 72(1):16, February 2020.
- [358] Andrea Petri, Zoltán Haiman, and Morgan May. Sample variance in weak lensing: How many simulations are required? *Phys. Rev. D*, 93(6):063524, March 2016.
- [359] Ryuichi Takahashi, Takashi Hamana, Masato Shirasaki, Toshiya Namikawa, et al. Full-sky Gravitational Lensing Simulation for Large-area Galaxy Surveys and Cosmic Microwave Background Experiments. *Astrophys. J.*, 850(1):24, November 2017.
- [360] Masato Shirasaki, Kana Moriwaki, Taira Oogi, Naoki Yoshida, et al. Noise reduction for weak lensing mass mapping: an application of generative adversarial networks to Subaru Hyper Suprime-Cam first-year data. *Mon. Not. Roy. Astron. Soc.*, 504(2):1825–1839, June 2021.
- [361] Masato Shirasaki and Naoki Yoshida. Statistical and Systematic Errors in the Measurement of Weak-Lensing Minkowski Functionals: Application to the Canada-France-Hawaii Lensing Survey. *Astrophys. J.*, 786(1):43, May 2014.
- [362] Masato Shirasaki, Masahiro Takada, Hironao Miyatake, Ryuichi Takahashi, et al. Robust covariance estimation of galaxy-galaxy weak lensing: validation and limitation of jackknife covariance. *Mon. Not. Roy. Astron. Soc.*, 470(3):3476–3496, September 2017.
- [363] Masato Shirasaki, Takashi Hamana, Masahiro Takada, Ryuichi Takahashi, et al. Mock galaxy shape catalogues in the Subaru Hyper Suprime-Cam Survey. *Mon. Not. Roy. Astron. Soc.*, 486(1):52–69, June 2019.
- [364] Alan F. Heavens, Raul Jimenez, and Ofer Lahav. Massive lossless data compression and multiple parameter estimation from galaxy spectra. *Mon. Not. Roy. Astron. Soc.*, 317(4):965–972, October 2000.
- [365] Cora Uhlemann, Olivier Friedrich, Aoife Boyle, Alex Gough, et al. It takes two to know one: Computing accurate one-point PDF covariances from effective two-point PDF models. *The Open Journal of Astrophysics*, 6:1, January 2023.
- [366] Daniel Foreman-Mackey, Alex Conley, Will Meierjorgen Farr, David W. Hogg, et al. emcee: The MCMC Hammer. *Astrophysics Source Code Library*, record ascl:1303.002, March 2013.

- [367] Daniel Foreman-Mackey, David W. Hogg, Dustin Lang, and Jonathan Goodman. emcee: The MCMC Hammer. *Publ. Astron. Soc. Pac.*, 125(925):306, March 2013.
- [368] Aki Vehtari, Andrew Gelman, Daniel Simpson, Bob Carpenter, et al. Rank-normalization, folding, and localization: An improved \hat{R} for assessing convergence of MCMC. *arXiv e-prints*, page arXiv:1903.08008, March 2019.
- [369] Rachel Mandelbaum, François Lanusse, Alexie Leauthaud, Robert Armstrong, et al. Weak lensing shear calibration with simulations of the HSC survey. *Mon. Not. Roy. Astron. Soc.*, 481(3):3170–3195, December 2018.
- [370] Ken Osato, Jia Liu, and Zoltán Haiman. κ TNG: effect of baryonic processes on weak lensing with IllustrisTNG simulations. *Mon. Not. Roy. Astron. Soc.*, 502(4):5593–5602, April 2021.
- [371] Joachim Harnois-Déraps, Nicolas Martinet, and Robert Reischke. Cosmic shear beyond 2-point statistics: Accounting for galaxy intrinsic alignment with projected tidal fields. *Mon. Not. Roy. Astron. Soc.*, 509(3):3868–3888, January 2022.
- [372] Xiangkun Liu, Shuo Yuan, Chuzhong Pan, Tianyu Zhang, et al. Cosmological studies from HSC-SSP tomographic weak-lensing peak abundances. *Mon. Not. Roy. Astron. Soc.*, 519(1):594–612, February 2023.
- [373] Leander Thiele, Elena Massara, Alice Pisani, ChangHoon Hahn, et al. Neutrino mass constraint from an Implicit Likelihood Analysis of BOSS voids. *arXiv e-prints*, page arXiv:2307.07555, July 2023.
- [374] John N. Bahcall and Raymond Davis, Jr. Solar Neutrinos: A Scientific Puzzle. *Science*, 191(4224):264–267, January 1976.
- [375] L. Wolfenstein. Neutrino oscillations in matter. *Phys. Rev. D*, 17(9):2369–2374, May 1978.
- [376] S. P. Mikheyev and A. Yu. Smirnov. Resonance enhancement of oscillations in matter and solar neutrino spectroscopy. *Yadernaya Fizika*, 42:1441–1448, January 1985.
- [377] Super-Kamiokande Collaboration, Y. Fukuda, et al. Evidence for Oscillation of Atmospheric Neutrinos. *Phys. Rev. Lett.*, 81(8):1562–1567, August 1998.
- [378] SNO Collaboration, Q. R. Ahmad, et al. Direct Evidence for Neutrino Flavor Transformation from Neutral-Current Interactions in the Sudbury Neutrino Observatory. *Phys. Rev. Lett.*, 89(1):011301, July 2002.
- [379] KamLAND Collaboration, T. Araki, et al. Measurement of Neutrino Oscillation with KamLAND: Evidence of Spectral Distortion. *Phys. Rev. Lett.*, 94(8):081801, March 2005.

- [380] K2K Collaboration, M. H. Ahn, et al. Measurement of neutrino oscillation by the K2K experiment. *Phys. Rev. D*, 74(7):072003, October 2006.
- [381] Daya Bay Collaboration, F. P. An, et al. Observation of Electron-Antineutrino Disappearance at Daya Bay. *Phys. Rev. Lett.*, 108(17):171803, April 2012.
- [382] KATRIN Collaboration, M. Aker, et al. KATRIN: status and prospects for the neutrino mass and beyond. *Journal of Physics G Nuclear Physics*, 49(10):100501, October 2022.
- [383] V. Icke. Voids and filaments. *Mon. Not. Roy. Astron. Soc.*, 206:1P–3P, January 1984.
- [384] Alice Pisani, Elena Massara, David N. Spergel, et al. Cosmic voids: a novel probe to shed light on our Universe. *Bull. Am. Astron. Soc.*, 51(3):40, May 2019.
- [385] Michele Moresco et al. Unveiling the Universe with emerging cosmological probes. *Living Reviews in Relativity*, 25(1):6, December 2022.
- [386] Nico Schuster, Nico Hamaus, Klaus Dolag, and Jochen Weller. Why cosmic voids matter: nonlinear structure & linear dynamics. *J. Cosmol. Astropart. Phys.*, 2023(5):031, May 2023.
- [387] Elena Massara, Francisco Villaescusa-Navarro, Matteo Viel, and P. M. Sutter. Voids in massive neutrino cosmologies. *J. Cosmol. Astropart. Phys.*, 2015(11):018–018, November 2015.
- [388] Arka Banerjee and Neal Dalal. Simulating nonlinear cosmological structure formation with massive neutrinos. *J. Cosmol. Astropart. Phys.*, 2016(11):015, November 2016.
- [389] Christina D. Kreisch, Alice Pisani, Carmelita Carbone, Jia Liu, et al. Massive neutrinos leave fingerprints on cosmic voids. *Mon. Not. Roy. Astron. Soc.*, 488(3):4413–4426, September 2019.
- [390] Nico Schuster, Nico Hamaus, Alice Pisani, Carmelita Carbone, et al. The bias of cosmic voids in the presence of massive neutrinos. *J. Cosmol. Astropart. Phys.*, 2019(12):055, December 2019.
- [391] Sofia Contarini, Federico Marulli, Lauro Moscardini, Alfonso Veropalumbo, et al. Cosmic voids in modified gravity models with massive neutrinos. *Mon. Not. Roy. Astron. Soc.*, 504(4):5021–5038, July 2021.
- [392] Giovanni Verza, Carmelita Carbone, Alice Pisani, and Alessandro Renzi. DEM-NUni: disentangling dark energy from massive neutrinos with the void size function. *arXiv e-prints*, page arXiv:2212.09740, December 2022.

- [393] Martin Sahlén. Cluster-void degeneracy breaking: Neutrino properties and dark energy. *Phys. Rev. D*, 99(6):063525, March 2019.
- [394] Adrian E. Bayer, Francisco Villaescusa-Navarro, Elena Massara, Jia Liu, et al. Detecting Neutrino Mass by Combining Matter Clustering, Halos, and Voids. *Astrophys. J.*, 919(1):24, September 2021.
- [395] Christina D. Kreisch, Alice Pisani, Francisco Villaescusa-Navarro, David N. Spergel, et al. The GIGANTES Data Set: Precision Cosmology from Voids in the Machine-learning Era. *Astrophys. J.*, 935(2):100, August 2022.
- [396] Selim C. Hotinli, Nashwan Sabti, Jaxon North, and Marc Kamionkowski. Unveiling Neutrino Halos with CMB Lensing. *arXiv e-prints*, page arXiv:2306.15715, June 2023.
- [397] S. A. Gregory and L. A. Thompson. The Coma/A1367 supercluster and its environs. *Astrophys. J.*, 222:784–799, June 1978.
- [398] Mikhel Jõeveer, Jaan Einasto, and Erik Tago. Spatial distribution of galaxies and of clusters of galaxies in the southern galactic hemisphere. *Mon. Not. Roy. Astron. Soc.*, 185:357–370, November 1978.
- [399] R. B. Tully and J. R. Fisher. Nearby Small Groups of Galaxies. In M. S. Longair and J. Einasto, editors, *Large Scale Structures in the Universe*, volume 79, page 31, January 1978.
- [400] R. P. Kirshner, Jr. Oemler, A., P. L. Schechter, et al. A million cubic megaparsec void in Bootes ? *Astrophys. J. Lett.*, 248:L57–L60, September 1981.
- [401] V. de Lapparent, M. J. Geller, and J. P. Huchra. A Slice of the Universe. *Astrophys. J. Lett.*, 302:L1, March 1986.
- [402] Martin Sahlén, Íñigo Zubeldía, and Joseph Silk. Cluster-Void Degeneracy Breaking: Dark Energy, Planck, and the Largest Cluster and Void. *Astrophys. J. Lett.*, 820(1):L7, March 2016.
- [403] Fiona Hoyle and Michael S. Vogeley. Voids in the Two-Degree Field Galaxy Redshift Survey. *Astrophys. J.*, 607(2):751–764, June 2004.
- [404] Danny C. Pan, Michael S. Vogeley, Fiona Hoyle, Yun-Young Choi, et al. Cosmic voids in Sloan Digital Sky Survey Data Release 7. *Mon. Not. Roy. Astron. Soc.*, 421(2):926–934, April 2012.
- [405] P. M. Sutter, Guilhem Lavaux, Benjamin D. Wandelt, and David H. Weinberg. A Public Void Catalog from the SDSS DR7 Galaxy Redshift Surveys Based on the Watershed Transform. *Astrophys. J.*, 761(1):44, December 2012.
- [406] P. M. Sutter, Guilhem Lavaux, Benjamin D. Wandelt, David H. Weinberg, et al. Voids in the SDSS DR9: observations, simulations, and the impact of the survey mask. *Mon. Not. Roy. Astron. Soc.*, 442(4):3127–3137, August 2014.

- [407] Seshadri Nadathur. Testing cosmology with a catalogue of voids in the BOSS galaxy surveys. *Mon. Not. Roy. Astron. Soc.*, 461(1):358–370, September 2016.
- [408] Qingqing Mao, Andreas A. Berlind, Robert J. Scherrer, Mark C. Neyrinck, et al. A Cosmic Void Catalog of SDSS DR12 BOSS Galaxies. *Astrophys. J.*, 835(2):161, February 2017.
- [409] P. M. Sutter, Guilhem Lavaux, Benjamin D. Wandelt, and David H. Weinberg. A First Application of the Alcock-Paczynski Test to Stacked Cosmic Voids. *Astrophys. J.*, 761(2):187, December 2012.
- [410] P. M. Sutter, Alice Pisani, Benjamin D. Wandelt, and David H. Weinberg. A measurement of the Alcock-Paczynski effect using cosmic voids in the SDSS. *Mon. Not. Roy. Astron. Soc.*, 443(4):2983–2990, October 2014.
- [411] Nico Hamaus, Alice Pisani, P. M. Sutter, Guilhem Lavaux, et al. Constraints on Cosmology and Gravity from the Dynamics of Voids. *Phys. Rev. Lett.*, 117(9):091302, August 2016.
- [412] Nico Hamaus, Marie-Claude Cousinou, Alice Pisani, Marie Aubert, et al. Multipole analysis of redshift-space distortions around cosmic voids. *J. Cosmol. Astropart. Phys.*, 2017(7):014, July 2017.
- [413] Qingqing Mao, Andreas A. Berlind, Robert J. Scherrer, Mark C. Neyrinck, et al. Cosmic Voids in the SDSS DR12 BOSS Galaxy Sample: The Alcock-Paczynski Test. *Astrophys. J.*, 835(2):160, February 2017.
- [414] Nico Hamaus, Alice Pisani, Jin-Ah Choi, Guilhem Lavaux, et al. Precision cosmology with voids in the final BOSS data. *J. Cosmol. Astropart. Phys.*, 2020(12):023, December 2020.
- [415] Seshadri Nadathur, Alex Woodfinden, Will J. Percival, et al. The completed SDSS-IV extended baryon oscillation spectroscopic survey: geometry and growth from the anisotropic void-galaxy correlation function in the luminous red galaxy sample. *Mon. Not. Roy. Astron. Soc.*, 499(3):4140–4157, December 2020.
- [416] Marie Aubert, Marie-Claude Cousinou, Stéphanie Escoffier, Adam J. Hawken, et al. The completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: growth rate of structure measurement from cosmic voids. *Mon. Not. Roy. Astron. Soc.*, 513(1):186–203, June 2022.
- [417] Alex Woodfinden, Will J. Percival, Seshadri Nadathur, Hans A. Winther, et al. Cosmological measurements from void-galaxy and galaxy-galaxy clustering in the Sloan Digital Sky Survey. *Mon. Not. Roy. Astron. Soc.*, June 2023.
- [418] Sofia Contarini, Alice Pisani, Nico Hamaus, Federico Marulli, et al. Cosmological constraints from the BOSS DR12 void size function. *arXiv e-prints*, page arXiv:2212.03873, December 2022.

- [419] Sofia Contarini, Alice Pisani, Nico Hamaus, Federico Marulli, et al. Voids fill us in on rising cosmology tensions. *arXiv e-prints*, page arXiv:2212.07438, December 2022.
- [420] SDSS Collaboration, Donald G. York, et al. The Sloan Digital Sky Survey: Technical Summary. *Astron. J.*, 120(3):1579–1587, September 2000.
- [421] SDSS Collaboration, Daniel J. Eisenstein, David H. Weinberg, et al. SDSS-III: Massive Spectroscopic Surveys of the Distant Universe, the Milky Way, and Extra-Solar Planetary Systems. *Astron. J.*, 142(3):72, September 2011.
- [422] SDSS Collaboration, Kyle S. Dawson, David J. Schlegel, et al. The Baryon Oscillation Spectroscopic Survey of SDSS-III. *Astron. J.*, 145(1):10, January 2013.
- [423] 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. *Phys. Rev. D*, 101(8):083504, April 2020.
- [424] Agne Semenaite, Ariel G. Sánchez, Andrea Pezzotta, Jiamin Hou, et al. Beyond Λ CDM constraints from the full shape clustering measurements from BOSS and eBOSS. *Mon. Not. Roy. Astron. Soc.*, 521(4):5013–5025, June 2023.
- [425] Stephen Stopyra, Hiranya V. Peiris, and Andrew Pontzen. How to build a catalogue of linearly evolving cosmic voids. *Mon. Not. Roy. Astron. Soc.*, 500(3):4173–4180, January 2021.
- [426] William H. Press and Paul Schechter. Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation. *Astrophys. J.*, 187:425–438, February 1974.
- [427] J. R. Bond, S. Cole, G. Efstathiou, and N. Kaiser. Excursion Set Mass Functions for Hierarchical Gaussian Fluctuations. *Astrophys. J.*, 379:440, October 1991.
- [428] Ravi K. Sheth and Giuseppe Tormen. Large-scale bias and the peak background split. *Mon. Not. Roy. Astron. Soc.*, 308(1):119–126, September 1999.
- [429] Ravi K. Sheth and Rien van de Weygaert. A hierarchy of voids: much ado about nothing. *Mon. Not. Roy. Astron. Soc.*, 350(2):517–538, May 2004.
- [430] Aseem Paranjape, Tsz Yan Lam, and Ravi K. Sheth. Halo abundances and counts-in-cells: the excursion set approach with correlated steps. *Mon. Not. Roy. Astron. Soc.*, 420(2):1429–1441, February 2012.
- [431] Aseem Paranjape, Ravi K. Sheth, and Vincent Desjacques. Excursion set peaks: a self-consistent model of dark halo abundances and clustering. *Mon. Not. Roy. Astron. Soc.*, 431(2):1503–1512, May 2013.

- [432] Elise Jennings, Yin Li, and Wayne Hu. The abundance of voids and the excursion set formalism. *Mon. Not. Roy. Astron. Soc.*, 434(3):2167–2181, September 2013.
- [433] Alice Pisani, P. M. Sutter, Nico Hamaus, Esfandiar Alizadeh, et al. Counting voids to probe dark energy. *Phys. Rev. D*, 92(8):083531, October 2015.
- [434] Tommaso Ronconi and Federico Marulli. Cosmological exploitation of cosmic void statistics. New numerical tools in the CosmoBolognaLib to extract cosmological constraints from the void size function. *Astron. Astrophys.*, 607:A24, October 2017.
- [435] T. Ronconi, S. Contarini, F. Marulli, M. Baldi, et al. Cosmic voids uncovered - first-order statistics of depressions in the biased density field. *Mon. Not. Roy. Astron. Soc.*, 488(4):5075–5084, October 2019.
- [436] Giovanni Verza, Alice Pisani, Carmelita Carbone, Nico Hamaus, et al. The void size function in dynamical dark energy cosmologies. *J. Cosmol. Astropart. Phys.*, 2019(12):040, December 2019.
- [437] Sofia Contarini, Tommaso Ronconi, Federico Marulli, Lauro Moscardini, et al. Cosmological exploitation of the size function of cosmic voids identified in the distribution of biased tracers. *Mon. Not. Roy. Astron. Soc.*, 488(3):3526–3540, September 2019.
- [438] S. Contarini, G. Verza, A. Pisani, N. Hamaus, et al. Euclid: Cosmological forecasts from the void size function. *Astron. Astrophys.*, 667:A162, November 2022.
- [439] D. Pelliciari, S. Contarini, F. Marulli, L. Moscardini, et al. Exploring the cosmological synergy between galaxy cluster and cosmic void number counts. *Mon. Not. Roy. Astron. Soc.*, 522(1):152–164, June 2023.
- [440] C. Alcock and B. Paczynski. An evolution free test for non-zero cosmological constant. *Nature*, 281:358, October 1979.
- [441] Guilhem Lavaux and Benjamin D. Wandelt. Precision Cosmography with Stacked Voids. *Astrophys. J.*, 754(2):109, August 2012.
- [442] A. Pisani, G. Lavaux, P. M. Sutter, and B. D. Wandelt. Real-space density profile reconstruction of stacked voids. *Mon. Not. Roy. Astron. Soc.*, 443(4):3238–3250, October 2014.
- [443] N. D. Padilla, L. Ceccarelli, and D. G. Lambas. Spatial and dynamical properties of voids in a Λ cold dark matter universe. *Mon. Not. Roy. Astron. Soc.*, 363(3):977–990, November 2005.
- [444] Dante Paz, Marcelo Lares, Laura Ceccarelli, Nelson Padilla, et al. Clues on void evolution-II. Measuring density and velocity profiles on SDSS galaxy redshift space distortions. *Mon. Not. Roy. Astron. Soc.*, 436(4):3480–3491, December 2013.

- [445] Nico Hamaus, P. M. Sutter, and Benjamin D. Wandelt. Modeling cosmic void statistics. *arXiv e-prints*, page arXiv:1409.7621, September 2014.
- [446] Nico Hamaus, P. M. Sutter, and Benjamin D. Wandelt. Universal Density Profile for Cosmic Voids. *Phys. Rev. Lett.*, 112(25):251302, June 2014.
- [447] Elena Massara and Ravi K. Sheth. Density and velocity profiles around cosmic voids. *arXiv e-prints*, page arXiv:1811.03132, November 2018.
- [448] Yu Feng, Man-Yat Chu, Uroš Seljak, and Patrick McDonald. FASTPM: a new scheme for fast simulations of dark matter and haloes. *Mon. Not. Roy. Astron. Soc.*, 463(3):2273–2286, December 2016.
- [449] Adrian E. Bayer, Arka Banerjee, and Yu Feng. A fast particle-mesh simulation of non-linear cosmological structure formation with massive neutrinos. *J. Cosmol. Astropart. Phys.*, 2021(1):016, January 2021.
- [450] Andreas A. Berlind and David H. Weinberg. The Halo Occupation Distribution: Toward an Empirical Determination of the Relation between Galaxies and Mass. *Astrophys. J.*, 575(2):587–616, August 2002.
- [451] Risa H. Wechsler and Jeremy L. Tinker. The Connection Between Galaxies and Their Dark Matter Halos. *Annu. Rev. Astron. Astrophys.*, 56:435–487, September 2018.
- [452] Kyle Cranmer, Johann Brehmer, and Gilles Louppe. The frontier of simulation-based inference. *Proceedings of the National Academy of Science*, 117(48):30055–30062, December 2020.
- [453] ChangHoon Hahn, Michael Eickenberg, Shirley Ho, Jiamin Hou, et al. SIMBIG: A Forward Modeling Approach To Analyzing Galaxy Clustering. *arXiv e-prints*, page arXiv:2211.00723, November 2022.
- [454] ChangHoon Hahn, Michael Eickenberg, Shirley Ho, Jiamin Hou, et al. SIMBIG: mock challenge for a forward modeling approach to galaxy clustering. *J. Cosmol. Astropart. Phys.*, 2023(4):010, April 2023.
- [455] Oliver H. E. Philcox and Mikhail M. Ivanov. BOSS DR12 full-shape cosmology: Λ CDM constraints from the large-scale galaxy power spectrum and bispectrum monopole. *Phys. Rev. D*, 105(4):043517, February 2022.
- [456] Antony Lewis, Anthony Challinor, and Anthony Lasenby. Efficient computation of CMB anisotropies in closed FRW models. *Astrophys. J.*, 538:473–476, 2000.
- [457] Cullan Howlett, Antony Lewis, Alex Hall, and Anthony Challinor. CMB power spectrum parameter degeneracies in the era of precision cosmology. *J. Cosmol. Astropart. Phys.*, 1204:027, 2012.

- [458] Diego Blas, Julien Lesgourgues, and Thomas Tram. The Cosmic Linear Anisotropy Solving System (CLASS). Part II: Approximation schemes. *J. Cosmol. Astropart. Phys.*, 2011(7):034, July 2011.
- [459] Matteo Zennaro, Julien Bel, Francisco Villaescusa-Navarro, Carmelita Carbone, et al. REPS: REscaled Power Spectra for initial conditions with massive neutrinos. Astrophysics Source Code Library, record ascl:1612.022, December 2016.
- [460] M. Zennaro, J. Bel, F. Villaescusa-Navarro, C. Carbone, et al. Initial conditions for accurate N-body simulations of massive neutrino cosmologies. *Mon. Not. Roy. Astron. Soc.*, 466(3):3244–3258, April 2017.
- [461] Andrei Variu, Shadab Alam, Cheng Zhao, Chia-Hsun Chuang, et al. DESI Mock Challenge: Constructing DESI galaxy catalogues based on FastPM simulations. *arXiv e-prints*, page arXiv:2307.14197, July 2023.
- [462] Francisco Villaescusa-Navarro, ChangHoon Hahn, Elena Massara, Arka Banerjee, et al. The Quijote Simulations. *Astrophys. J. Suppl.*, 250(1):2, September 2020.
- [463] Zheng Zheng, Andreas A. Berlind, David H. Weinberg, Andrew J. Benson, et al. Theoretical Models of the Halo Occupation Distribution: Separating Central and Satellite Galaxies. *Astrophys. J.*, 633(2):791–809, November 2005.
- [464] Guangtun Zhu, Zheng Zheng, W. P. Lin, Y. P. Jing, et al. The Dependence of the Occupation of Galaxies on the Halo Formation Time. *Astrophys. J. Lett.*, 639(1):L5–L8, March 2006.
- [465] Andrew R. Zentner, Andrew P. Hearin, and Frank C. van den Bosch. Galaxy assembly bias: a significant source of systematic error in the galaxy-halo relationship. *Mon. Not. Roy. Astron. Soc.*, 443(4):3044–3067, October 2014.
- [466] Arnau Pujol and Enrique Gaztañaga. Are the halo occupation predictions consistent with large-scale galaxy clustering? *Mon. Not. Roy. Astron. Soc.*, 442(3):1930–1941, August 2014.
- [467] Beth A. Reid, Hee-Jong Seo, Alexie Leauthaud, Jeremy L. Tinker, et al. A 2.5 per cent measurement of the growth rate from small-scale redshift space clustering of SDSS-III CMASS galaxies. *Mon. Not. Roy. Astron. Soc.*, 444(1):476–502, October 2014.
- [468] Yen-Ting Lin, Rachel Mandelbaum, Yun-Hsin Huang, Hung-Jin Huang, et al. On Detecting Halo Assembly Bias with Galaxy Populations. *Astrophys. J.*, 819(2):119, March 2016.
- [469] Yosuke Kobayashi, Takahiro Nishimichi, Masahiro Takada, and Hironao Miyatake. Full-shape cosmology analysis of the SDSS-III BOSS galaxy power spectrum using an emulator-based halo model: A 5% determination of σ_8 . *Phys. Rev. D*, 105(8):083517, April 2022.

- [470] Andreas A. Berlind, David H. Weinberg, Andrew J. Benson, Carlton M. Baugh, et al. The Halo Occupation Distribution and the Physics of Galaxy Formation. *Astrophys. J.*, 593(1):1–25, August 2003.
- [471] Kohji Yoshikawa, Y. P. Jing, and Gerhard Börner. Spatial and Dynamical Biases in Velocity Statistics of Galaxies. *Astrophys. J.*, 590(2):654–663, June 2003.
- [472] Frank C. van den Bosch, Simone M. Weinmann, Xiaohu Yang, H. J. Mo, et al. The phase-space parameters of the brightest halo galaxies. *Mon. Not. Roy. Astron. Soc.*, 361(4):1203–1215, August 2005.
- [473] Hong Guo, Zheng Zheng, Idit Zehavi, Kyle Dawson, et al. Velocity bias from the small-scale clustering of SDSS-III BOSS galaxies. *Mon. Not. Roy. Astron. Soc.*, 446(1):578–594, January 2015.
- [474] Zhongxu Zhai, Jeremy L. Tinker, Arka Banerjee, Joseph DeRose, et al. The Aemulus Project. V. Cosmological Constraint from Small-scale Clustering of BOSS Galaxies. *Astrophys. J.*, 948(2):99, May 2023.
- [475] Jordan Carlson and Martin White. Embedding Realistic Surveys in Simulations Through Volume Remapping. *Astrophys. J. Suppl.*, 190(2):311–314, October 2010.
- [476] P. M. Sutter, G. Lavaux, N. Hamaus, A. Pisani, et al. VIDE: The Void IDentification and Examination toolkit. *Astronomy and Computing*, 9:1–9, March 2015.
- [477] Mark C. Neyrinck. ZOBOV: a parameter-free void-finding algorithm. *Mon. Not. Roy. Astron. Soc.*, 386(4):2101–2109, June 2008.
- [478] Jörg M. Colberg, Frazer Pearce, Caroline Foster, Erwin Platen, et al. The Aspen-Amsterdam void finder comparison project. *Mon. Not. Roy. Astron. Soc.*, 387(2):933–944, June 2008.
- [479] R. van de Weygaert and W. Schaap. The Cosmic Web: Geometric Analysis. In V. J. Martínez, E. Saar, E. Martínez-González, and M. J. Pons-Bordería, editors, *Data Analysis in Cosmology*, volume 665, pages 291–413. 2009.
- [480] Marius Cautun, Enrique Paillas, Yan-Chuan Cai, Sownak Bose, et al. The Santiago-Harvard-Edinburgh-Durham void comparison - I. SHEDding light on chameleon gravity tests. *Mon. Not. Roy. Astron. Soc.*, 476(3):3195–3217, May 2018.
- [481] Nick Hand, Yu Feng, Florian Beutler, Yin Li, et al. nbodykit: An Open-source, Massively Parallel Toolkit for Large-scale Structure. *Astron. J.*, 156(4):160, October 2018.

- [482] Nick Hand, Yu Feng, Florian Beutler, Yin Li, et al. nbodykit: Massively parallel, large-scale structure toolkit. *Astrophysics Source Code Library*, record ascl:1904.027, April 2019.
- [483] Hume A. Feldman, Nick Kaiser, and John A. Peacock. Power-Spectrum Analysis of Three-dimensional Redshift Surveys. *Astrophys. J.*, 426:23, May 1994.
- [484] Beth Reid, Shirley Ho, Nikhil Padmanabhan, Will J. Percival, et al. SDSS-III Baryon Oscillation Spectroscopic Survey Data Release 12: galaxy target selection and large-scale structure catalogues. *Mon. Not. Roy. Astron. Soc.*, 455(2):1553–1573, January 2016.
- [485] M. C. Cousinou, A. Pisani, A. Tilquin, N. Hamaus, et al. Multivariate analysis of cosmic void characteristics. *Astronomy and Computing*, 27:53, April 2019.
- [486] Alice Pisani, P. M. Sutter, and B. D. Wandelt. Mastering the effects of peculiar velocities in cosmic voids. *arXiv e-prints*, page arXiv:1506.07982, June 2015.
- [487] Mikhail M. Ivanov, Marko Simonović, and Matias Zaldarriaga. Cosmological parameters from the BOSS galaxy power spectrum. *J. Cosmol. Astropart. Phys.*, 2020(5):042, May 2020.
- [488] Kyle Cranmer, Juan Pavez, and Gilles Louppe. Approximating Likelihood Ratios with Calibrated Discriminative Classifiers. *arXiv e-prints*, page arXiv:1506.02169, June 2015.
- [489] Joeri Hermans, Volodimir Begy, and Gilles Louppe. Likelihood-free MCMC with Amortized Approximate Ratio Estimators. *arXiv e-prints*, page arXiv:1903.04057, March 2019.
- [490] Conor Durkan, Iain Murray, and George Papamakarios. On Contrastive Learning for Likelihood-free Inference. *arXiv e-prints*, page arXiv:2002.03712, February 2020.
- [491] Arnaud Delaunoy, Joeri Hermans, François Rozet, Antoine Wehenkel, et al. Towards Reliable Simulation-Based Inference with Balanced Neural Ratio Estimation. *arXiv e-prints*, page arXiv:2208.13624, August 2022.
- [492] Benjamin Kurt Miller, Christoph Weniger, and Patrick Forré. Contrastive Neural Ratio Estimation. *arXiv e-prints*, page arXiv:2210.06170, October 2022.
- [493] Alvaro Tejero-Cantero, Jan Boelts, Michael Deistler, Jan-Matthis Lueckmann, et al. sbi: A toolkit for simulation-based inference. *Journal of Open Source Software*, 5(52):2505, 2020.
- [494] Justin Alsing and Benjamin Wandelt. Nuisance hardened data compression for fast likelihood-free inference. *Mon. Not. Roy. Astron. Soc.*, 488(4):5093–5103, October 2019.

- [495] Guido d'Amico, Jérôme Gleyzes, Nickolas Kokron, Katarina Markovic, et al. The cosmological analysis of the SDSS/BOSS data from the Effective Field Theory of Large-Scale Structure. *J. Cosmol. Astropart. Phys.*, 2020(5):005, May 2020.
- [496] Shi-Fan Chen, Zvonimir Vlah, and Martin White. A new analysis of galaxy 2-point functions in the BOSS survey, including full-shape information and post-reconstruction BAO. *J. Cosmol. Astropart. Phys.*, 2022(2):008, February 2022.
- [497] Oliver H. E. Philcox, Mikhail M. Ivanov, Marko Simonović, and Matias Zaldarriaga. Combining full-shape and BAO analyses of galaxy power spectra: a 1.6% CMB-independent constraint on H_0 . *J. Cosmol. Astropart. Phys.*, 2020(5):032, May 2020.
- [498] Oliver H. E. Philcox. Cosmology without window functions: Quadratic estimators for the galaxy power spectrum. *Phys. Rev. D*, 103(10):103504, May 2021.
- [499] Anton Chudaykin, Mikhail M. Ivanov, Oliver H. E. Philcox, and Marko Simonović. Nonlinear perturbation theory extension of the Boltzmann code CLASS. *Phys. Rev. D*, 102(6):063533, September 2020.
- [500] Joeri Hermans, Arnaud Delaunoy, François Rozet, Antoine Wehenkel, et al. A Trust Crisis In Simulation-Based Inference? Your Posterior Approximations Can Be Unfaithful. *arXiv e-prints*, page arXiv:2110.06581, October 2021.
- [501] Francis J. Anscombe. The transformation of Poisson, binomial and negative-binomial data. *Biometrika*, 35(3-4):246–254, 12 1948.
- [502] Arjun Dey, David J. Schlegel, Dustin Lang, et al. Overview of the DESI Legacy Imaging Surveys. *Astron. J.*, 157(5):168, May 2019.
- [503] Olivier Doré, Jamie Bock, et al. Cosmology with the SPHEREX All-Sky Spectral Survey. *arXiv e-prints*, page arXiv:1412.4872, December 2014.
- [504] Masahiro Takada, Richard S. Ellis, Masashi Chiba, Jenny E. Greene, et al. Extragalactic science, cosmology, and Galactic archaeology with the Subaru Prime Focus Spectrograph. *Publ. Astron. Soc. Jpn.*, 66(1):R1, February 2014.
- [505] J. Chluba, E. Switzer, K. Nelson, and D. Nagai. Sunyaev-Zeldovich signal processing and temperature-velocity moment method for individual clusters. *Mon. Not. Roy. Astron. Soc.*, 430:3054–3069, April 2013.
- [506] W. Hu, D. Scott, and J. Silk. Reionization and cosmic microwave background distortions: A complete treatment of second-order Compton scattering. *Phys. Rev. D*, 49:648–670, January 1994.

- [507] N. Battaglia, H. Trac, R. Cen, and A. Loeb. Reionization on Large Scales. I. A Parametric Model Constructed from Radiation-hydrodynamic Simulations. *Astrophys. J.*, 776(2):81, October 2013.
- [508] Digvijay Wadekar, Leander Thiele, Francisco Villaescusa-Navarro, J. Colin Hill, et al. Augmenting astrophysical scaling relations with machine learning : application to reducing the SZ flux-mass scatter. *arXiv e-prints*, page arXiv:2201.01305, January 2022.
- [509] Michael D. Schmidt and Hod Lipson. Distilling free-form natural laws from experimental data. *Science*, 324:81 – 85, 2009.
- [510] Digvijay Wadekar, Francisco Villaescusa-Navarro, Shirley Ho, and Laurence Perreault-Levasseur. Modeling assembly bias with machine learning and symbolic regression. *arXiv e-prints*, page arXiv:2012.00111, November 2020.
- [511] Miles Cranmer, Alvaro Sanchez-Gonzalez, Peter Battaglia, Rui Xu, et al. Discovering Symbolic Models from Deep Learning with Inductive Biases. *arXiv e-prints*, page arXiv:2006.11287, June 2020.
- [512] Miles D. Cranmer, Rui Xu, Peter Battaglia, and Shirley Ho. Learning Symbolic Physics with Graph Networks. *arXiv e-prints*, page arXiv:1909.05862, September 2019.
- [513] Helen Shao, Francisco Villaescusa-Navarro, Shy Genel, David N. Spergel, et al. Finding universal relations in subhalo properties with artificial intelligence. *arXiv e-prints*, page arXiv:2109.04484, September 2021.
- [514] Silviu-Marian Udrescu and Max Tegmark. AI Feynman: A physics-inspired method for symbolic regression. *Science Advances*, 6(16):eaay2631, April 2020.
- [515] Tailin Wu and Max Tegmark. Toward an AI Physicist for Unsupervised Learning. *arXiv e-prints*, page arXiv:1810.10525, October 2018.
- [516] Samuel Kim, Peter Y. Lu, Sriyon Mukherjee, Michael Gilbert, et al. Integration of Neural Network-Based Symbolic Regression in Deep Learning for Scientific Discovery. *arXiv e-prints*, page arXiv:1912.04825, December 2019.
- [517] Ziming Liu and Max Tegmark. AI Poincaré: Machine Learning Conservation Laws from Trajectories. *arXiv e-prints*, page arXiv:2011.04698, November 2020.
- [518] Casper Wilstrup and Jaan Kasak. Symbolic regression outperforms other models for small data sets. *arXiv e-prints*, page arXiv:2103.15147, March 2021.
- [519] Severin Strobl, Arno Formella, and Thorsten Pöschel. Exact calculation of the overlap volume of spheres and mesh elements. *Journal of Computational Physics*, 311:158 – 172, 2016.

- [520] A. S. G. Robotham and Cullan Howlett. A Short Research Note on Calculating Exact Distribution Functions and Random Sampling for the 3D NFW Profile. *Research Notes of the American Astronomical Society*, 2(2):55, June 2018.
- [521] Andrew P. Hearin, Andrew R. Zentner, Frank C. van den Bosch, Duncan Campbell, et al. Introducing decorated HODs: modelling assembly bias in the galaxy-halo connection. *Mon. Not. Roy. Astron. Soc.*, 460(3):2552–2570, August 2016.
- [522] Andrew Hearin, Erik Tollerud, Thomas Robitaille, Michael Droettboom, et al. Halotools: Galaxy-Halo connection models. Astrophysics Source Code Library, record ascl:1604.005, April 2016.

Appendix A

Appendix to Chapter 2

A.1 Assessment of prior volume effects

The more extended parameterizations considered in this paper, $f_2 = \dots$, exhibit a pathology in that $f_1 = 0$ is equivalent to Λ CDM regardless of the value of Δ_1 . Thus, there may be a prior volume effect in these models that underestimates the posterior for non-vanishing baryon clumping (and thus increased H_0). We assess this effect for the example of the $f_2 = 1/2$ model; for other choices of f_2 the conclusions are similar. Our methods are similar to those used in Ref. [20].

First, we consider changing the model such that a constraint $f_1 > 0.1$ is enforced, thereby avoiding Λ CDM entirely. The resulting posteriors are shown in Fig. A.1. It is expected that the posteriors for b and H_0 move upward to some extent since the model, by construction, contains at least some baryon clumping now. The shift in the central value of H_0 is about $\sigma/3$ and we do not observe a substantial enlargement of the high- H_0 tail.

Second, we compute a frequentist mean likelihood profile in b , which is the exponential of the average log-likelihood extracted from the Markov chain in 20 b -bins between 0 and 1, shown as the solid blue line in Fig. A.2. We observe that this profile

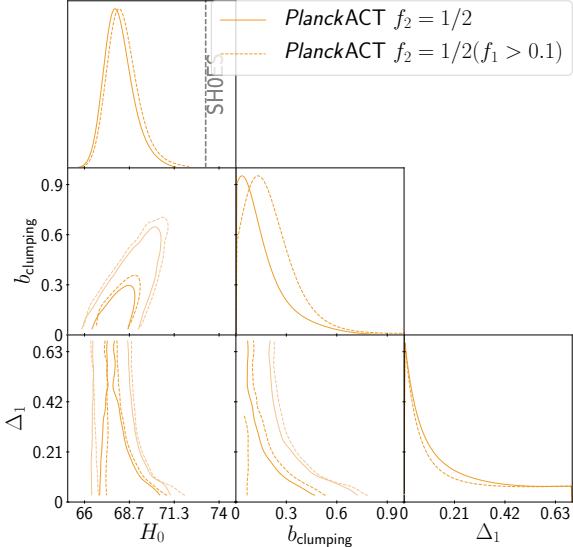


Figure A.1: Comparison between the posterior in the unconstrained $f_2 = 1/2$ model, same as in Fig. 2.3, and the one with the constraint $f_1 > 0.1$. The latter should substantially reduce the prior volume effect arising from the degeneracy with ΛCDM at $f_1 = 0$. In the dashed posterior, the non-vanishing support at $b = 0$ is an artifact of `MontePython`'s posterior smoothing routine.

is somewhat broader than the posterior shown in Fig. 2.3. Furthermore, at the 95 % upper limit on b from Tab. A.1, the averaged likelihood is reduced by about $\Delta\chi^2 \sim 2.4$. This value would correspond to only an $\sim 88\%$ upper bound.

We also compute the more common likelihood profile in which we minimize the likelihood at a range of values of b , shown as the solid green line in Fig. A.2. In this case, we observe that the maximum is at non-zero b , but then the likelihood profile drops off more steeply than the marginalized Bayesian posterior. For a discussion of the relative merits of the two ways to compute a likelihood profile, see Ref. [20].

We conclude that we observe evidence for a mild prior volume effect. However, the posteriors in Fig. A.1 indicate that our main conclusions regarding H_0 are only marginally affected.

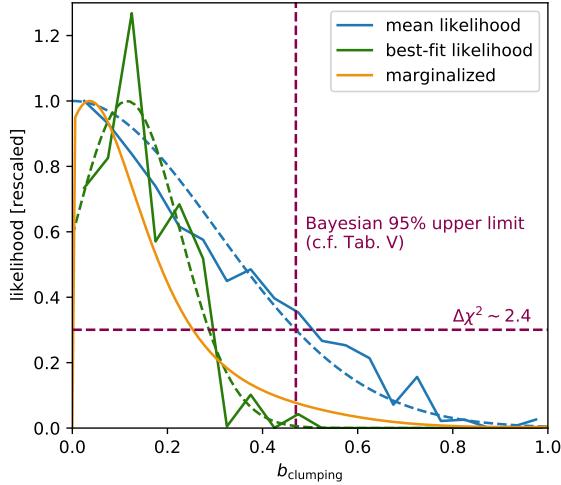


Figure A.2: Likelihood profiles as a function of b for the $f_2 = 1/2$ Markov chain. The ‘mean’ and ‘best-fit’ lines are computed as described in the text. The orange line is the same marginalized likelihood that was shown in Fig. 2.3. The blue dashed line is a Gaussian fit with only the standard deviation varied, while the green dashed line is a 3-parameter Gaussian fit. The green solid line was rescaled so that the Gaussian fit attains its maximum at one. These fits should only be treated as visual guides.

A.2 Central values and uncertainties

likelihood	model	H_0 [km/s/Mpc]	$100\omega_b$	ω_{cdm}	n_s	$10^9 A_s$	τ_{reio}	b
<i>Planck</i>	ΛCDM	67.24 ± 0.66	2.23(2)	0.120(1)	0.964(5)	2.10(4)	0.054(9)	—
	M1	68.45 ± 1.08	2.24(2)	0.122(2)	0.960(5)	2.10(4)	0.053(8)	<0.57
	M2	67.86 ± 0.83	2.24(2)	0.121(2)	0.964(5)	2.10(4)	0.055(9)	<0.32
	$f_2 = 1/3$	68.04 ± 0.97	2.24(2)	0.121(2)	0.963(6)	2.10(4)	0.054(9)	<0.42
	$f_2 = 1/2$	68.12 ± 1.04	2.24(2)	0.121(2)	0.963(6)	2.10(4)	0.054(9)	<0.53
	$f_2 = 2/3$	68.03 ± 0.96	2.24(2)	0.121(2)	0.964(6)	2.10(4)	0.054(9)	<0.52
<i>Planck+ACT</i>	ΛCDM	67.26 ± 0.60	2.23(1)	0.120(1)	0.967(4)	2.12(4)	0.055(8)	—
	M1	68.18 ± 0.87	2.23(1)	0.121(2)	0.962(5)	2.11(4)	0.053(8)	<0.42
	M2	67.74 ± 0.71	2.24(1)	0.121(1)	0.967(4)	2.12(4)	0.054(8)	<0.26
	$f_2 = 1/3$	67.99 ± 0.86	2.23(1)	0.121(2)	0.964(6)	2.11(4)	0.054(8)	<0.39
	$f_2 = 1/2$	68.06 ± 0.88	2.23(1)	0.121(2)	0.964(6)	2.11(4)	0.054(8)	<0.47
	$f_2 = 2/3$	68.01 ± 0.89	2.24(2)	0.121(2)	0.965(5)	2.11(4)	0.054(8)	<0.49
<i>Planck+ACT+BAO</i>	$f_2 = 1/3$	68.18 ± 0.77	2.24(2)	0.121(2)	0.964(6)	2.11(4)	0.054(8)	<0.38
	$f_2 = 1/2$	68.24 ± 0.83	2.24(1)	0.121(2)	0.964(6)	2.11(4)	0.054(8)	<0.46
	$f_2 = 2/3$	68.22 ± 0.79	2.24(1)	0.121(2)	0.966(5)	2.11(4)	0.054(9)	<0.50

Table A.1: Central values and $1-\sigma$ errors for the posteriors in this section. For conciseness, for ΛCDM parameters except H_0 , the numbers in parentheses are the $1-\sigma$ uncertainties on the last digit. Since we do not observe a detection of the clumping effect, for the clumping parameter b 95 % upper limits are quoted.

Appendix B

Appendix to Chapter 3

B.1 Electron temperature and the relativistic SZ effect

A cluster's thermal SZ contribution for a given line of sight can be written as

$$\Delta I_\nu(\hat{n}) = I_0 \int N_e(\hat{n}, l) \sigma_T S_\nu(T_e(\hat{n}, l)) dl \quad (\text{B.1})$$

where $I_0 = B_\nu(T_0)$ is the CMB blackbody spectrum, N_e denotes the electron number density, dl parameterizes the line of sight integration and $S_\nu(T)$ determines the SZ spectrum with relativistic temperature corrections. If the electron temperature is constant along the line of sight, $T_e(\hat{n}, l) \equiv T_e(\hat{n})$, one can simply write $\Delta I_\nu(\hat{n}) = \tau(\hat{n}) I_0 S_\nu(T_e(\hat{n}))$, where the Thomson optical depth is $\tau(\hat{n}) = \int N_e(\hat{n}, l) \sigma_T dl$. However, generally the temperature varies along the line of sight, such that a moment expansion provides a simpler method for analyzing and describing the SZ signal [505].

Defining $S_\nu^{(k)}(T) \equiv \partial^k S_\nu(T)/\partial T^k$, we can perform the moment expansion of the SZ signal around the τ -weighted temperature, $T_e^\tau(\hat{n}) = \int N_e(\hat{n}, l) \sigma_T T_e(\hat{n}, l) dl / \tau(\hat{n})$.

Up to second order in temperature, this yields

$$\Delta I_\nu(\hat{n}) \approx \tau I_0 \left\{ S_\nu(T_e^\tau) + \frac{1}{2} S_\nu^{(2)}(T_e^\tau) [\langle T_e^2 \rangle_\tau - (T_e^\tau)^2] \right\}, \quad (\text{B.2})$$

where we suppressed the dependence on \hat{n} and introduced the τ -weighted temperature moments

$$\langle T_e^k \rangle_\tau = \frac{\int N_e(\hat{n}, l) \sigma_T T_e^k(\hat{n}, l) dl}{\tau(\hat{n})} \quad (\text{B.3})$$

such that $T_e^\tau \equiv \langle T_e \rangle_\tau$. For the standard Λ CDM cosmology the first two temperature moments from halos alone are $\langle T_e \rangle_\tau = 0.208 \text{ keV}$ and $\langle T_e^2 \rangle_\tau = 0.299 \text{ keV}^2$ [85].

Alternatively, we can perform the moment expansion around the y -weighted temperature, T_e^y . By introducing the y -weighted moments

$$\langle T_e^k \rangle_y = \frac{\int N_e(\hat{n}, l) \sigma_T \frac{T_e(\hat{n}, l)}{m_e} T_e^k(\hat{n}, l) dl}{y(\hat{n})} \quad (\text{B.4})$$

with $y(\hat{n}) = \int N_e(\hat{n}, l) \sigma_T \frac{T_e(\hat{n}, l)}{m_e} dl$ and $T_e^y \equiv \langle T_e \rangle_y$, again to second order in the temperature one then has

$$\Delta I_\nu(\hat{n}) \approx y I_0 \left\{ \tilde{S}_\nu(T_e^y) + \frac{1}{2} \tilde{S}_\nu^{(2)}(T_e^y) [\langle T_e^2 \rangle_y - (T_e^y)^2] \right\}. \quad (\text{B.5})$$

where $\tilde{S}_\nu = S_\nu/\Theta_e$, $\tilde{S}_\nu^{(k)} = \partial^k \tilde{S}_\nu(T)/\partial T^k$, and we introduced the dimensionless temperature $\Theta_e = \frac{T_e}{m_e}$. These two representations are essentially equivalent¹; however, the latter is slightly more economic when it comes to capturing the relativistic tSZ effect, as we will see next.

Low temperature limit and mix of hot and cold gas: In the low temperature limit one can write [e.g., 79, 81]

$$S_\nu(T) \approx \Theta_e Y_0(x) + \Theta_e^2 Y_1(x) \quad (\text{B.6})$$

¹They become indistinguishable when more temperature terms are included in the expansion.

where $x = \frac{h\nu}{T_0}$ and the functions Y_0 and Y_1 are the first two terms of the asymptotic expansion for the tSZ signal [e.g., see 81]. It is clear that in this limit, only two independent spectral parameters can be determined. This is directly evident when using the y -weighted moments with $\tilde{S}_\nu^{(2)}(T) \approx 0$:

$$\frac{\Delta I_\nu(\hat{n})}{I_0} \approx y S_\nu(T_e^y) = y \left[Y_0(x) + \frac{T_e^y}{m_e} Y_1(x) \right]. \quad (\text{B.7})$$

Since $T_e^y \equiv \langle T_e^2 \rangle_\tau / \langle T_e \rangle_\tau$, the two important parameters for the spectral analysis in the Λ CDM case are expected to be $y \approx 1.77 \times 10^{-6}$ and $T_e^y \approx 1.44 \text{ keV}$ [85]. However, the presence of extremely cold gas from reionization modifies the observational inference. This can be seen by adding another y -distortion contribution with no relativistic correction

$$\frac{\Delta I_\nu^{\text{tot}}(\hat{n})}{I_0} \approx y^{\text{re}} Y_0(x) + y \left[Y_0(x) + \frac{T_e^y}{m_e} Y_1(x) \right]. \quad (\text{B.8})$$

The reionization y -parameter, y^{re} , is roughly 10% of the cluster contribution [506, 85] and cannot be distinguished from the cluster contribution with distortion measurements alone. In an analysis, an effective electron temperature of $T_e^{y,*} = T_e^y / (1 + y^{\text{re}}/y) \approx 1.30 \text{ keV}$ would thus be recovered. Due to small contributions from higher order temperature corrections, the recovered result for Λ CDM is $T_e^{y,*} \approx 1.24 \text{ keV}$ [78].

B.2 Cosmology dependence

Using the halo model we compute how the distortion monopoles $\langle y \rangle$ and $\langle T_e \rangle$ depend on the relevant Λ CDM parameters. We use the halo model fitting formulae described in Sec. 3.4.2 to compute the observables $x_i = \{\langle y \rangle, \langle T_e \rangle\}$ as functions of $\theta_j = \{h, \Omega_m, \Omega_b, n_s, \sigma_8\}$. The resulting functions are well approximated by power laws,

$x_i = x_i^{(0)} (\theta_j / \theta_j^{(0)})^{\alpha_{ij}}$. The resulting one-parameter fits are given by:

$$\begin{aligned}
10^6 \langle y \rangle(h) &= 1.48(h/0.6711)^{1.59}, \\
\text{keV}^{-1} \langle T_e \rangle(h) &= 1.51(h/0.6711)^{-0.86}, \\
10^6 \langle y \rangle(\Omega_m) &= 1.48(\Omega_m/0.3)^{0.92}, \\
\text{keV}^{-1} \langle T_e \rangle(\Omega_m) &= 1.51(\Omega_m/0.3)^{0.36}, \\
10^6 \langle y \rangle(\Omega_b) &= 1.48(\Omega_b/0.0490)^{0.84}, \\
\text{keV}^{-1} \langle T_e \rangle(\Omega_b) &= 1.51(\Omega_b/0.0490)^{0.09}, \\
10^6 \langle y \rangle(n_s) &= 1.48(n_s/0.9624)^{1.52}, \\
\text{keV}^{-1} \langle T_e \rangle(n_s) &= 1.51(n_s/0.9624)^{-0.84}, \\
10^6 \langle y \rangle(\sigma_8) &= 1.48(\sigma_8/0.8)^{3.86}, \\
\text{keV}^{-1} \langle T_e \rangle(\sigma_8) &= 1.51(\sigma_8/0.8)^{1.93}.
\end{aligned} \tag{B.9}$$

Note that the given scalings have a number of uncertainties. First, they only include the halo contribution. Second, the assumed cluster temperature model may not be optimal, as discussed in Sec. 3.4.2. Thus, we caution against blind use of these equations. However, for small deviations from the pivot cosmology they should provide reasonable approximations for the slope, even though the prefactors are most likely not very useful.

B.3 Reionization contribution

The non-relativistic $\langle y \rangle$ receives a small contribution from the epoch of reionization, which is not modelled in our simulations. This appendix describes an estimate of the magnitude and uncertainty of this effect. We calculate the $\langle y \rangle$ from reionization using maps constructed by ray-tracing through the past light cone of a semi-analytic realization from $z = 5.5$ to $z = 20$, which defines the redshift range we consider for

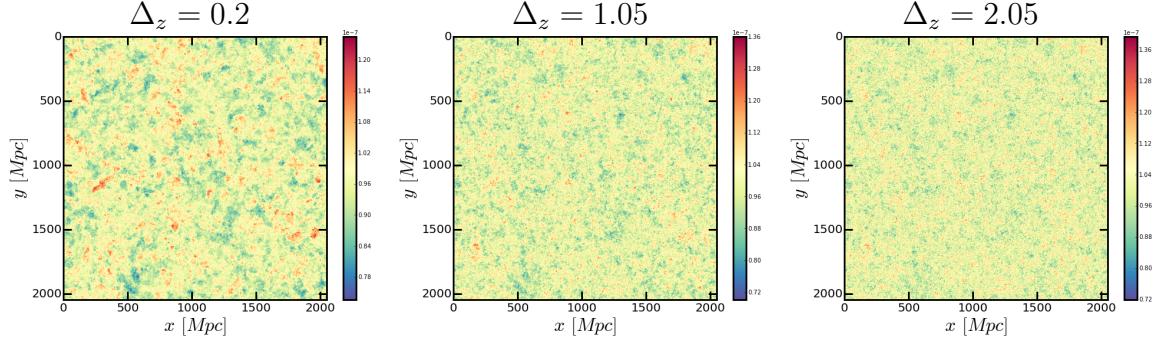


Figure B.1: Maps of the Compton- y field from reionization, generated using a semi-analytic model on a gravity-only simulation. The panels assume different durations of reionization, with the short, fiducial, and long duration models from left to right. Note the slightly different color scales.

reionization. In this semi-analytic realization, reionization fields are constructed on a gravity-only simulation following the method in Ref. [507]. For each spatial cell in the N -body simulation we have a density and ionization state as a function of time. We set the initial temperature of the cells when they reionize as $T_0 = 2 \times 10^4$ K, then the i th cell cools adiabatically according to

$$\frac{T_i(z)}{T_0} = \frac{1+z}{1+z_{i,\text{RE}}}, \quad (\text{B.10})$$

where $z_{i,\text{RE}}$ is the redshift at which the given cell reionizes. Using the parametric reionization model from Ref. [507], we change the duration and midpoint of reionization from the fiducial parameters. We show Compton- y maps from reionization for our fiducial model and the two extreme duration models, short and long duration, in Fig. B.1. For the fiducial reionization model, with a median reionization redshift of $z_{\text{mid}} = 10$, where 50 % of the Universe has reionized by mass, and a duration parameter $\Delta_z = 1.05$, we find that $\langle y \rangle_{\text{reio}} = 9.9 \times 10^{-8}$. For the short ($\Delta_z = 0.2$) and long ($\Delta_z = 2.05$) duration models we find 9.6×10^{-8} and 1.0×10^{-7} , respectively, while the dependence on z_{mid} (which we varied in [8, 12]) is smaller. Thus, reionization

contributes less than 10% to the total signal and we estimate the uncertainty as $\sim 5 \times 10^{-9}$, significantly below the error budget for our assumed experiment.

Clearly, the relativistic distortion receives minuscule contributions from reionization as $T_e/m_e \lesssim 10^{-4}$.

B.4 IGM contribution

In this appendix we compute the contribution from the intergalactic medium (IGM) to the $\langle y \rangle$ signal. We define this quantity as all Compton- y generated at $z < 5.5$ (which marks the end of reionization, c.f. Appendix B.3) outside of any halo. Since the IGM pressure is low compared to the ICM, a reasonable approximation is

$$\langle y \rangle_{\text{IGM}} = \frac{\sigma_T}{m_e} \int dl \bar{n}_e(z) T_{\text{IGM}}(z) \quad (\text{B.11})$$

where the integration is over distance up to $z = 5.5$, $\bar{n}_e(z) \equiv x_e(z)\Omega_b\rho_{\text{crit}}(z)$, and $x_e(z)$ is the free electron fraction. In our fiducial model, we assume $T_{\text{IGM},0} = 2 \times 10^4$ K for the IGM temperature at the end of reionization. We then assume the temperature drops adiabatically. The function $x_e(z)$ will depend on the redshift z_{HeII} at which HeII reionizes. In our fiducial model we assume instantaneous HeII reionization at $z_{\text{HeII}} = 3.5$ such that

$$x_e(z) = \begin{cases} \frac{3X_H+1}{4}; & z > z_{\text{HeII}}, \\ \frac{X_H+1}{2}; & z < z_{\text{HeII}}, \end{cases} \quad (\text{B.12})$$

where X_H is the primordial hydrogen mass fraction. The fiducial model yields $\langle y \rangle_{\text{reio}} = 7.1 \times 10^{-8}$. Changes in the starting value of T_{IGM} linearly affect the Compton- y . Assuming $z_{\text{HeII}} = 2.5$ results in 7×10^{-8} and $z_{\text{HeII}} = 4.5$ gives 7.2×10^{-8} . Thus, the IGM contribution is in magnitude comparable to the reionization signal with somewhat larger theoretical uncertainty driven by the temperature normalization.

B.5 Two-parameter dependence

In this appendix, we consider dependence on pairs of parameters. The primary goal of this exercise is to establish how strongly couplings between feedback parameters affect the y observables. In Fig. B.2, in each panel we plot the quantity

$$\frac{x_i(A_j, A_k)}{\text{const} \times x_i(A_j)x_i(A_k)} - 1, \quad (\text{B.13})$$

where the constant normalizes such that the ratio is one at the fiducial point and the x_i are evaluations of the neural nets with either one or two parameters varied from the fiducial point. Thus, we illustrate deviations from perfect factorization. We observe that for IllustrisTNG the factorization is a rather good approximation, with couplings of at most 5 %. SIMBA exhibits stronger corrections, but the factorization is still relatively accurate to within $\sim 40\%$ and of course the variations are also much larger by about an order of magnitude. In units of the overall differences in the y observables the inter-parameters couplings are quite similar in IllustrisTNG and SIMBA. It must be noted that the edges of parameter space are likely not quite accurately represented by the neural nets. We should emphasize that the neural nets have no structural preference for factorized representations. In summary, most of the dependence of y observables on feedback parameters can be readily read off from Fig. 3.4.

B.6 Symbolic regression

Symbolic regression identifies equations with parsimonious combinations of input parameters that have the smallest scatter with the given quantity of interest [508, 509, 131, 510–518]. We employ it to predict $\langle y \rangle$ and $\langle T_e \rangle$ separately as a function of the

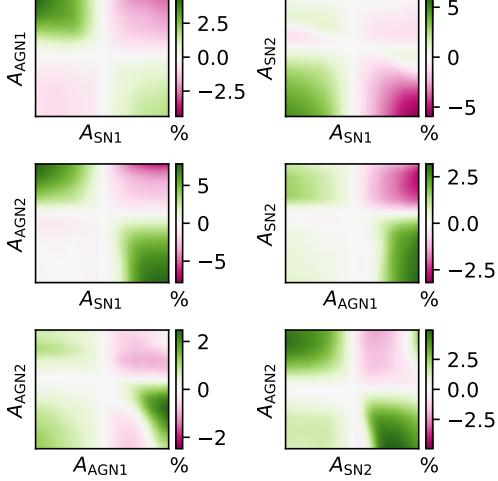
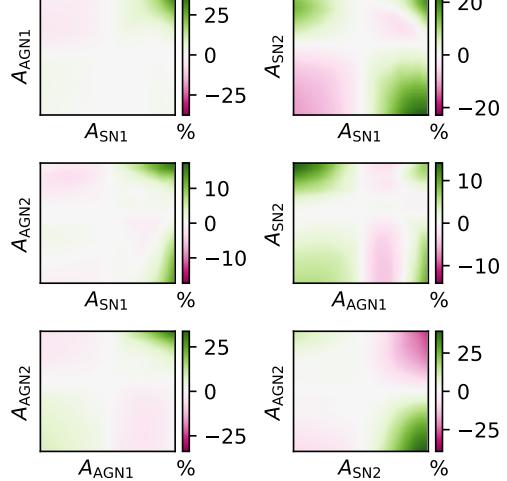
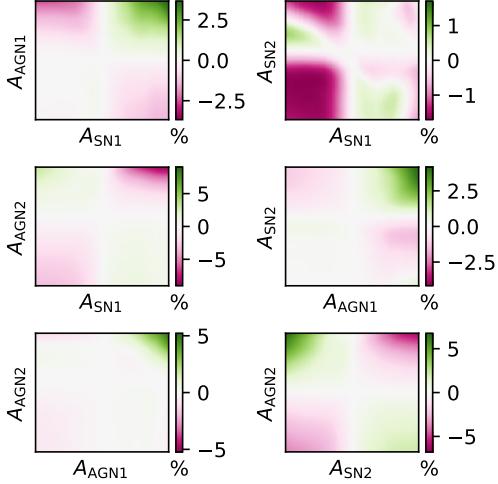
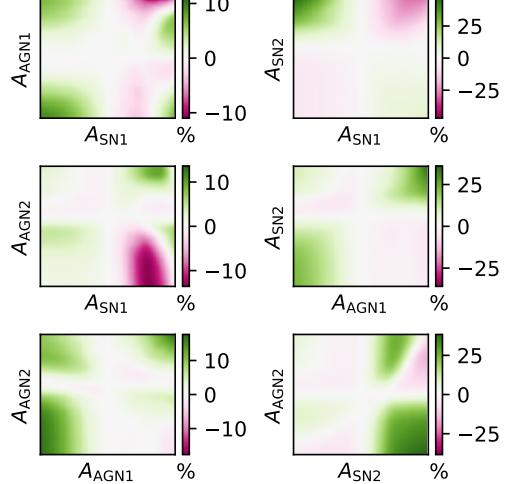
IllustrisTNG $\langle y \rangle$ SIMBA $\langle y \rangle$ IllustrisTNG $\langle T_e \rangle$ SIMBA $\langle T_e \rangle$ 

Figure B.2: Dependence of spectral distortions on pairs of simulation feedback parameters. Each panel shows evaluations of the neural networks trained on the LH set for two astrophysical parameters varied, divided by the simple model in which the dependence factorizes (using the curves from Fig. 3.4). As can be seen, for IllustrisTNG the factorization approximation is good to within $\sim 5\%$, while for SIMBA it is somewhat worse but still reasonable. For readability, the axes ticks have been suppressed. They are exactly identical to the ones in Fig. 3.4, so that the fiducial model with all feedback parameters equal one is in the center of each panel.

feedback parameters:

$$\{\langle y \rangle, \langle T_e \rangle\} = f(A_{\text{SN1}}, A_{\text{AGN1}}, A_{\text{SN2}}, A_{\text{AGN2}}) \quad (\text{B.14})$$

We show the equations obtained and their performance in Fig. B.3. The data points shown are the measurements from the CAMELS LH set (for the case of TNG, the data has been corrected for sample variance and the cosmology dependence has also been removed, c.f. Fig. 3.3). We also obtained equations more complex than the ones in Fig. B.3, however, as the risk of overfitting goes up as the equations get more complex, we show simple ones which have a substantial reduction in the mean squared error. The equations perform much better for the case of SIMBA. This could be either because the dependence of $\langle y \rangle, \langle T_e \rangle$ on the feedback parameters is weaker for TNG, or the dependence for the case of TNG might have a very complicated functional form and it is hard to find a good approximation given the limited size of our data set. Overall, the results in Fig. B.3 are consistent with the assumption in Eq. 3.12 that the individual parameter feedback dependence can be factorized.

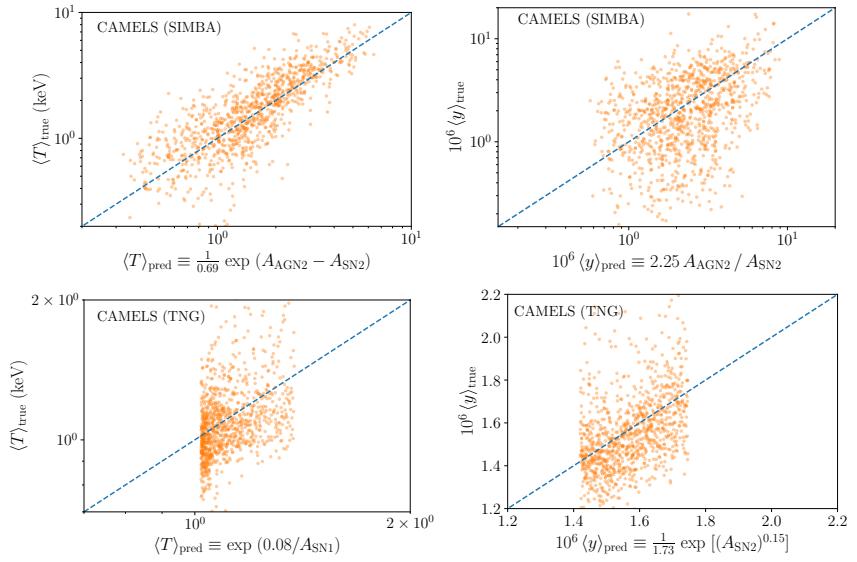


Figure B.3: As a complement to the results from the neural network, we use a machine learning tool called symbolic regression to predict $\langle y \rangle$ and $\langle T_e \rangle$ separately as an analytic function of the feedback parameters. The blue dashed line represents predicted=true; the closer the points are to the line the more accurate the prediction. The equations perform comparatively better for SIMBA. It is also interesting to see that the effects of AGN and supernovae on both $\langle y \rangle$ and $\langle T_e \rangle$ are opposite to each other for the case of SIMBA.

Appendix C

Appendix to Chapter 4

C.1 Pixelization

Our primary input data are properties of the dark matter field from the gravity-only simulations. For prediction of the electron pressure and density, we use the dark matter density, while for the electron momentum density we also use the dark matter momentum density. We make use of the `subfind_hsm1` fields in the snapshot to obtain spheres of varying size in which we assume the dark matter properties to be uniform; then we use the `Overlap` library¹ [519] to grid these spheres onto cubical voxels. The target data (electron pressure, density, and momentum density) are taken from the hydrodynamical run. Here, we approximate the Voronoi cells as spheres, the radius being inferred from the volume given in the simulation snapshot, and perform the gridding to voxels as described. These gridding procedures are approximations whose accuracy improves as the number of simulation elements in the individual voxels increases. Thus, these approximations are not worrisome in the high-density regions we are primarily focused on in this work. We note that the electron pressure is not directly measured in the simulation; however, under the assumption of complete ionization a linear relationship between thermal and electron pressure allows us to

¹<https://github.com/severinstrobl/overlap>

infer the latter. For reference, $P_{\text{th}}/P_e = (5X_{\text{H}} + 3)/2(X_{\text{H}} + 1) = 1.932$, with X_{H} the primordial hydrogen mass fraction.

C.2 Training boxes

As justified in Section 4.3.2, we restrict the target volumes to cubical boxes of size $(3.2 h^{-1}\text{Mpc})^3$; the semi-analytical models are evaluated on the same volume. The dark matter samples serving as input are taken as boxes of twice the sidelength, i.e. volume $(6.4 h^{-1}\text{Mpc})^3$; the target volume is centred in these boxes. The additional padding of $1.6 h^{-1}\text{Mpc}$ on each side of the target volume allows for inclusion of long-range effects. For electron pressure and density, we use voxels of sidelength $100.1 h^{-1}\text{kpc}$, such that the dark matter (electron gas) boxes are of size 64^3 (32^3) voxels. For the electron momentum density we use half this resolution (voxel sidelength $200.2 h^{-1}\text{kpc}$). We chose the resolution by two considerations: (1) we do not expect structure below length-scales of a few $100 h^{-1}\text{kpc}$ to be important in the mapping from dark matter to electron gas, while observationally smaller scales should also be of minor importance; (2) higher resolutions would be difficult to implement on the hardware available to us, the limiting factor is GPU memory.

For electron density and momentum density as targets, we split the TNG300 box into training (70%), validation (20%), and testing (10%) cuboidal sub-boxes. For the electron pressure we use external data for training (as discussed in Section 4.4.1), allowing us to use the whole simulation box for testing (we also use 20% of the box for validation, leading to an overlap between validation and testing data; however, the small volume in comparison to the testing box ensures that this does not invalidate the testing results).

C.3 Transformations

The numerical values a , b , c introduced in Eq. (4.11) are as follows:

	target = P_e	target = ρ_e	target = \mathbf{p}_e
	$x_{\text{DM}} = \rho_m$	$x_{\text{DM}} = \rho_m$	$x_{\text{DM}} = \rho_m$
a	2.2939	2.2160	8.4592
b	2.7984×10^6	2.7984×10^6	3.9381×10^6
c	3.8477×10^{-1}	3.8232×10^{-1}	2.6845×10^{-2}
			0.0000

We remind the reader of the unit conventions given in footnote 7. The calculations for a , b , and c were performed by restricting to “interesting” voxels in the sense of Section 4.4.1, these definitions were slightly different for different targets which explains the different values.

Appendix D

Appendix to Chapter 5

D.1 Qualitative comparison to CNN

In this section, we provide a qualitative comparison to the convolutional neural net (CNN) results from [156]. As we have argued in the main text, the set-based architecture is better able to structurally mirror the distinct features of the problem, namely sparseness and rotational equivariance.

Showing that these advantages also translate into better predictions compared to those obtained with CNNs in [156] is a non-trivial task, however. In that earlier work, the emphasis was on predicting an entire volume, while here we are constraining ourselves to massive halos (we have argued, however, that the remainder of the task should be relatively straightforward and can probably be solved with CNNs). Thus, [156] optimized the network for summary statistics (mostly the power spectrum) and worked at a fixed spatial resolution (while we, in this work, choose to adapt the resolution to individual halos, in the geometric spirit of the architecture).

Due to these caveats, a comprehensive quantitative comparison is beyond the scope of this work (and should probably be performed at the level of summary statistics one would eventually be interested in, which necessitates inclusion of redshift dependence).

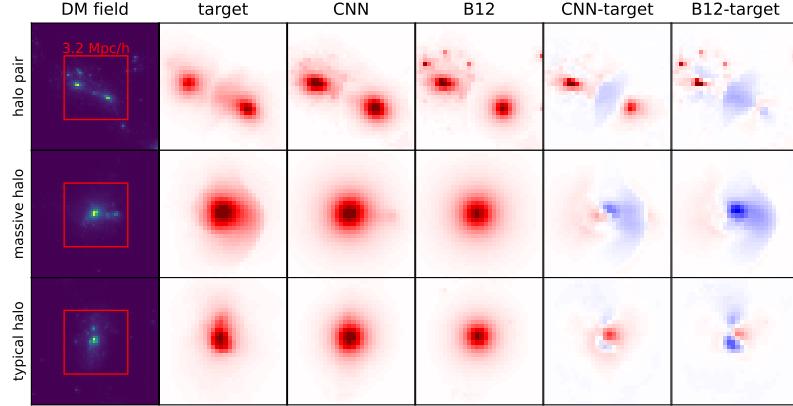


Figure D.1: For comparison with Fig. 5.3, we provide images of some halos obtained with the CNN from [156]. See the text for cautionary notes on direct comparison.

However, a relatively straightforward comparison can be made at the level of halo images. We have already presented integrated images of a few halos, obtained with our set-based network, in Fig. 5.3. For comparison, Fig. D.1 provides similar images obtained with the CNN (this figure is a modified version of Fig. 10 in [156]). The color scales are constructed identically to the way chosen in Fig. 5.3. We need to mention a few points in which the two figures are not directly comparable. First, the CNN images are at a fixed resolution, with $3.2 h^{-1}\text{Mpc}$ sidelength. Second, the semi-analytic model labeled ‘B12’ is similar to the GNFW model used as a benchmark in this work, but it was not refit to the specific simulation (which is IllustrisTNG-300, same as in this work). Only an overall scaling was applied to the ‘B12’ pressure profiles.

Despite these differences, at least a qualitative comparison is possible. The object labeled ‘massive halo’ is actually the same object as the right-most one in Fig. 5.3, with the same projection direction (it is a lucky coincidence that this object ended up in the testing set in [156], too). Here, we observe that the CNN tends to make the same mistakes as the B12 model, albeit to a lesser extent. This strong reliance on the semi-analytic model is a general trend observed with the CNN, while the set-based architecture proves to be relatively robust to mistakes made by the GNFW model. At

a more quantitative level, the set-based architecture presented in this work produces a more accurate prediction than the CNN, as evidenced by the darker residuals in Fig. D.1 compared to Fig. 5.3.

While for the other two objects in D.1 we unfortunately do not have counterparts in Fig. 5.3, the same trends appear to hold: the CNN is overly reliant on the B12 model, which translates into larger errors than those made by the set-based architecture.

Appendix E

Appendix to Chapter 6

E.1 Information

In Fig. E.1, we illustrate where the information in the PDF is primarily coming from.

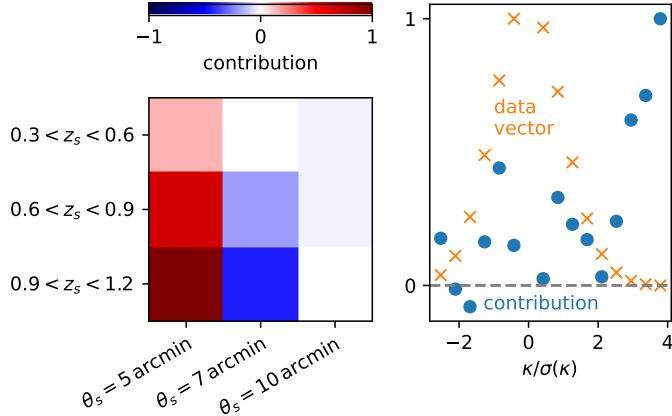


Figure E.1: Illustration of the compression weights for the PDF. We picked a high- S_8 cosmology (with close to fiducial Ω_m) from our simulations. By taking the element-wise product between the first MOPED compression vector (corresponding to S_8) and the data vector we can get intuition as to where information is coming from. For this figure, we subtract the corresponding product for the fiducial simulations. Note that the data vector shown in orange is just for illustration and has been rescaled and shifted for convenience. *Left:* contributions from different source redshifts and smoothing scales (with the κ -bins summed over). *Right:* contributions from different κ -bins with source redshifts and smoothing scales summed over.

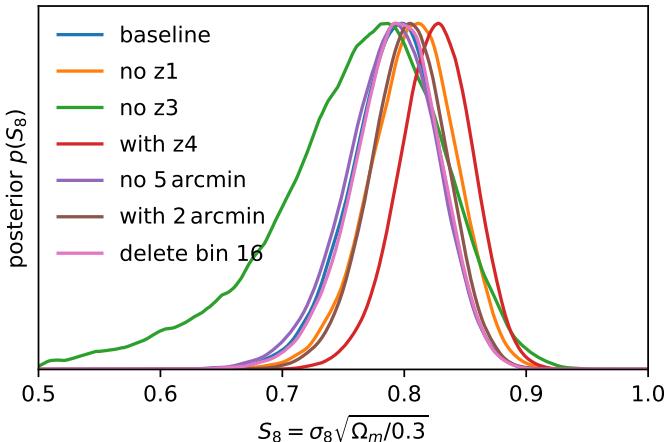


Figure E.2: The effect of various cuts and extensions in the data vector on the final posterior. All posteriors plotted are for $\text{PDF} + C_\ell^{\kappa\kappa}$, the blue baseline being the same curve as in Fig. 6.4.

In the left panel, we focus on redshift bins and smoothing scales. We observe that, analogously to the two-point statistics, the highest redshift bin is the most informative. Furthermore, the largest smoothing scale (10 arcmin) is quite uninformative, due to the nearly Gaussian character of the field. On the other hand, the smaller smoothing scales contain information and the compression appears to use a difference between them.

In the right panel, we focus on the convergence bins. Even though it might be expected that the higher signal-to-noise ratio close to the PDF's peak would make these bins the most informative, we observe that the high- κ tail is actually where most of the information is coming from. This is consistent with our picture in Fig. 6.1 that the deviations from Gaussianity in the tail are the focus of this work.

E.2 Data vector choices

In Fig. E.2, we explore the effects of various modifications to the data vector. First, we remove the lowest and highest tomographic redshift bin (orange, green) and observe that the higher one contributes much more signal-to-noise. Then, we add the tomographic redshift bin that was excluded in our baseline analysis, c.f. Sec. 6.2.2 (red). As discussed in the text, including this bin leads to a substantial shift in the

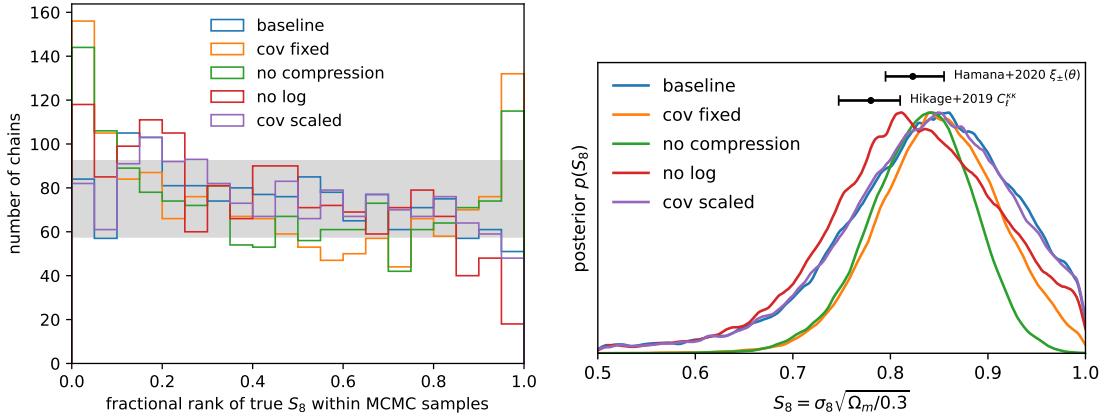


Figure E.3: The effect of changing the PDF likelihood in various ways (c.f. Sec. 6.2.4). In all plots, the data vector considered is PDF only. The left panel is the ranks plot, c.f. Fig. 6.2, and the right panel are the posteriors on real data obtained with the different likelihoods. We observe that the baseline likelihood used in the main part of the paper exhibits the healthiest calibration diagnostics. Furthermore, we see that the different likelihood assumptions lead to visible but mild shifts in the posterior.

posterior.

The last three tests apply only to the PDF part of the data vector. First, we remove the smallest smoothing scale (purple). Then we add a 2 arcmin smoothing scale (brown). We observe that using the smaller smoothing scale tightens the posterior substantially, indicating that there is useful information. However, in our baseline analysis such small scales are not used due to concerns about baryonic systematics. Finally, we delete the 17th convergence bin (instead of the 10th, c.f. Sec. 6.2.2), and find that this modification of the data vector does not alter the posterior.

E.3 Likelihood choices

In Fig. E.3, we show what happens to the PDF-only posterior when various choices in our likelihood construction are altered. As shown on the left panel, the baseline used in the main text (blue) exhibits the best rank statistics (closest to uniform distribution).

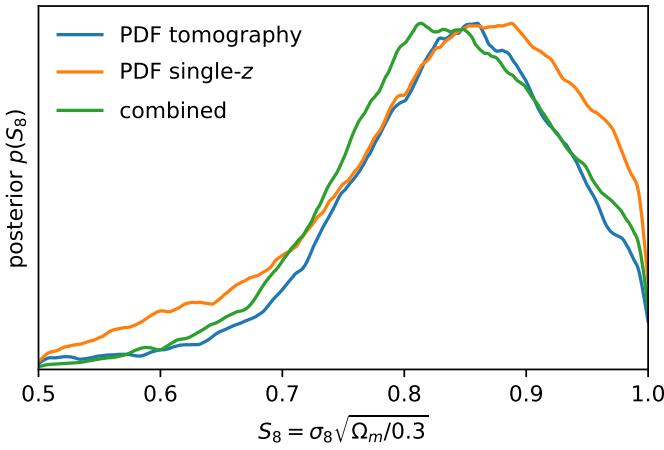


Figure E.4: Impact of tomography in the PDF-only analysis. We consider only the lower three redshift bins, as explained in Sec. 6.2.2. We observe that the posterior obtained with tomography, blue, is tighter than the one without, orange. The combined posterior, green, is consistent with the individual ones but slightly broader than the tomographic result, presumably due to small inefficiencies in the compression.

The modifications deviate from uniform distribution. The resulting posteriors are still broadly consistent.

E.4 Tomography

For two-point analyses, including the auto- and cross-power spectra from tomographic redshift bins saturates the information. The situation is more complicated with non-Gaussian statistics because the shape noise propagates in a non-linear way into the data vector and posteriors. In the main text, we used a tomographic analysis, motivated by previous work that indicated benefits to tomography [340]. In Fig. E.4, we show what happens when we collapse all sources in a single bin (orange). We observe that the tomographic analysis does yield a somewhat tighter posterior. Of course, one can also combine tomographic and non-tomographic data vectors, or use various other mergers of redshift bins. We do not expect such strategies to significantly affect the constraining power in our case, but more generally they warrant further study.

Appendix F

Appendix to Chapter 7

F.1 Halo occupation distribution

In this appendix, we provide a more detailed description of the adopted HOD model.

First, for reference, the baseline five-parameter model only depends on halo mass M and has mean occupations

$$\overline{N}_{\text{cen}} = \frac{1}{2} \left[1 + \text{erf} \left(\frac{\log M - \log M_{\min}}{\sigma_{\log M}} \right) \right] \quad (\text{F.1})$$

for the central galaxy and

$$\overline{N}_{\text{sat}} = \overline{N}_{\text{cen}} \left(\frac{M - M_0}{M_1} \right)^{\alpha} \quad (\text{F.2})$$

for the satellites. A central is placed with probability $\overline{N}_{\text{cen}}$ and the number of satellites is drawn from a Poisson distribution with mean $\overline{N}_{\text{sat}}$. The central is placed at the halo's center and assigned the halo velocity. The satellites are distributed isotropically with an NFW profile [225] and the concentration model of Ref. [224], using the analytic solution for the inverse NFW CDF from Ref. [520]. Satellite velocities are drawn from a distribution assuming virialization.

On top of this baseline model, we implement galaxy assembly bias using the decorated HOD [521, 522] with the ratio of kinetic to potential energy $r \equiv T/U$ as proxy for assembly history. In our preliminary tests T/U outperformed halo concentration, possibly due to limited resolution within the **FastPM** halos. The decoration works by splitting halos into two groups according to r . In order to reduce the effect of any potential evolution of r with halo mass, we do this splitting separately within 64 groups containing equal numbers of halos. The fraction P_1 of halos with lowest r is assigned type 1 (2) for positive (negative) a_{bias} , while the rest is assigned type 2 (1). Then, the mean occupations are modified as

$$\Delta \bar{N}_{\text{cen}} = |a_{\text{bias}}| \min \left[1 - \bar{N}_{\text{cen}}, \frac{1 - P_1}{P_1} \bar{N}_{\text{cen}} \right] \quad (\text{F.3})$$

$$\Delta \bar{N}_{\text{sat}} = |a_{\text{bias}}| \frac{1 - P_1}{P_1} \bar{N}_{\text{sat}} \quad (\text{F.4})$$

for type 1 and

$$\Delta \bar{N}_{\text{cen}} = |a_{\text{bias}}| \max \left[-\bar{N}_{\text{cen}}, \frac{1 - P_1}{P_1} (\bar{N}_{\text{cen}} - 1) \right] \quad (\text{F.5})$$

$$\Delta \bar{N}_{\text{sat}} = |a_{\text{bias}}| (-\bar{N}_{\text{sat}}) \quad (\text{F.6})$$

for type 2.

Velocity bias for the centrals is implemented by adding $\eta_{\text{cen}} V_{\text{vir}} n$ where $n \sim N(0, 1)$. For the satellites, the velocity difference from the host halo is scaled by η_{sat} .

Redshift dependence for M_{min} , M_1 is approximated as linear in scale factor, such that

$$\Delta \log M_i = \mu(M_i)(a - a_0) \quad (\text{F.7})$$

with $a_0 = 1/(1 + 0.53)$.

We adopt flat priors $12.5 < \log M_{\text{min}} < 13.2$, $0.1 < \sigma_{\log M} < 0.8$, $12.5 < \log M_{0,1} < 15.5$, $0.2 < \alpha < 1.5$, $-3 < P'_1 < 3$, $-1 < a_{\text{bias}} < 1$, $5 < \eta'_{\text{cen}} < 10$, $-1 < \eta'_{\text{sat}} < 1$,

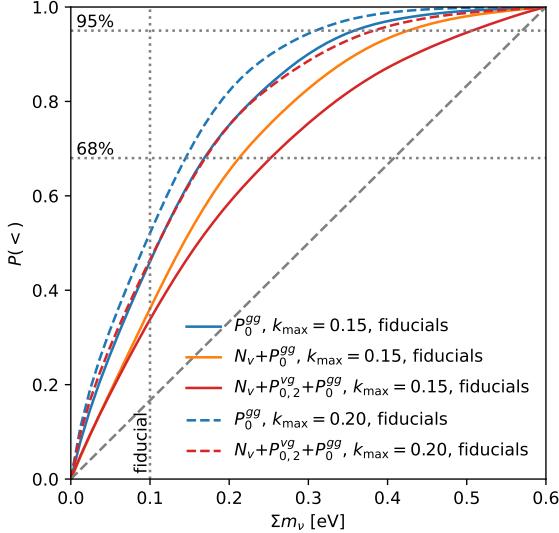


Figure F.1: Averaged posteriors on ~ 20 randomly chosen fiducial mocks. A widening of the $\sum m_\nu$ posterior upon including void statistics is also observed in simulations.

$-20 < \mu(M_{\min}) < 20$, $-40 < \mu(M_1) < 40$. Here, all masses are in $h^{-1}M_\odot$, and the primed parameters are defined as $2P_1 = (1 + \tanh P'_1)$, $\eta_{\text{sat}} = \exp(\eta'_{\text{sat}})$, $\eta_{\text{cen}} = \exp(-10 + \eta'_{\text{cen}})$. The above intervals were found during preliminary inference runs. Note in particular the small values of M_{\min} , compared to other analyses. This is partly explained by systematically lower halo masses in `FastPM`, and partly by the $n(z)$ downsampling described below. We picked the transformations given by the primed parameters based on the intuition that strictly (mathematically) bounded intervals often indicate that a uniform prior in a transformed quantity is a better choice.

F.2 Broadening of posteriors

We have seen in our main posteriors that adding void statistics to the data vector sometimes broadens the posterior on $\sum m_\nu$. In this appendix, we attempt to better understand this observation, focusing on $k_{\max} = 0.20$. For this, we will rely on inference on our fiducial mocks.

The first possible explanation could be a statistical fluctuation, and we cannot definitely exclude this hypothesis. One way, however, to test it is to look at *average* posteriors on our fiducial mocks. We perform inference on ~ 20 randomly chosen

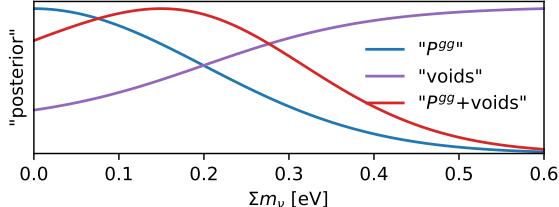


Figure F.2: Schematic illustration of the proposed mechanism explaining the broadening of posteriors when voids are included. This figure is only meant as a guide and not as a literal depiction of posteriors. Correlations and nuisance parameters complicate the picture.

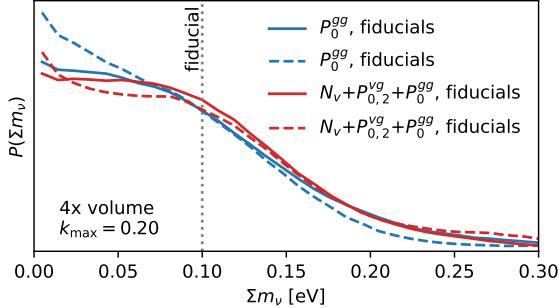


Figure F.3: Posteriors on joint analyses of four randomly chosen fiducial mocks, averaged over ~ 30 groups. The solid and dashed lines correspond to likelihoods with two different sets of five nuisance parameters kept explicit. We see that the posteriors where void statistics are included have a slightly more pronounced bump at the true value $\sum m_\nu = 0.1$ eV, consistent with the speculative picture in Fig. F.2.

mocks and plot the CDFs of concatenated chains in Fig. F.1. There, we observe that the expected, average behavior is for the posterior to broaden once void statistics are added to the data vector.

The second possible explanation could be that once void statistics are added our compression procedure becomes less efficient. This could certainly be the case if at linear order the void statistics appear more constraining than they are globally, thus P^{gg} would be unnecessarily downweighted. This hypothesis appears unlikely in light of the full posteriors presented in Figs. F.4, F.5. In these posteriors, we observe that for the parameters that are actually constrained (like M_{\min}) adding void statistics generically tightens the posteriors. It appears unlikely that $\sum m_\nu$ should be an exception.

Having found these two hypotheses unsatisfactory, we arrive at the third one: *void statistics tend to constrain $\sum m_\nu$ from below*. We illustrate this theory qualitatively in Fig. F.2, which should not be interpreted as a literal depiction. In fact, in Sec. 7.4.4 we show that void statistics alone yield posteriors close to the prior. Fig. F.2 provides merely an effective depiction.

We can investigate this hypothesis further by performing the following test. In order to increase signal-to-noise, we perform inference on four fiducial mocks at the same time, shown in Fig. F.3. For this, we use a different set of neural nets in which we leave five nuisance parameters explicit. The reason is that all implicitly marginalized nuisance parameters are effectively assumed to be different for each of the four mocks, an effect we would like to minimize. Of course, increasing the number of explicit nuisance parameters complicates the training and we have less confidence in the precise calibration of the posteriors. For this reason, our baseline results were obtained with only two explicit nuisance parameters. For reference, the real data posteriors obtained with these alternative neural nets are shown in Fig. F.5. We perform this test with two different sets of nuisance parameters kept explicit in order to gauge robustness (corresponding to the solid and dashed lines in Fig. F.3). Similar to Fig. F.1, we average posteriors over ~ 30 randomly chosen groups of four mocks in order to decrease sample variance. We observe that, consistent with our theory, the posteriors that include void statistics show a more pronounced hint of a bump at the true $\sum m_\nu$. In principle, one could increase the simulated volume further by combining more mocks, but our neural nets are not calibrated at the required level of precision and thus the resulting posteriors would not be robust.

In summary, the more mundane ideas to explain the observed broadening of posteriors appear questionable given the tests presented. On the other hand, the idea that void statistics are most effective at constraining $\sum m_\nu$ from below receives support from our experiments. A more in-depth examination of this issue would constitute a great starting point for future work.

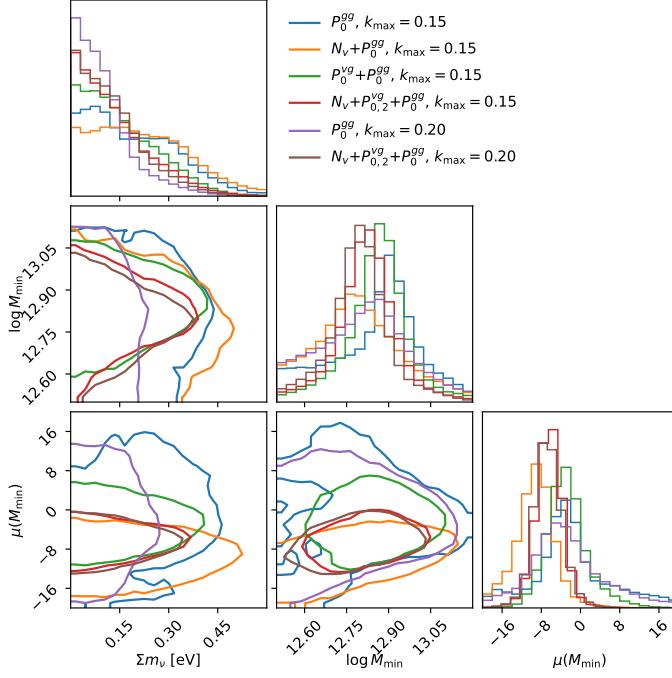


Figure F.4: Posteriors in the full parameter space considered, including the two HOD parameters we choose to keep explicit. The HOD posteriors are consistent between different analysis choices.

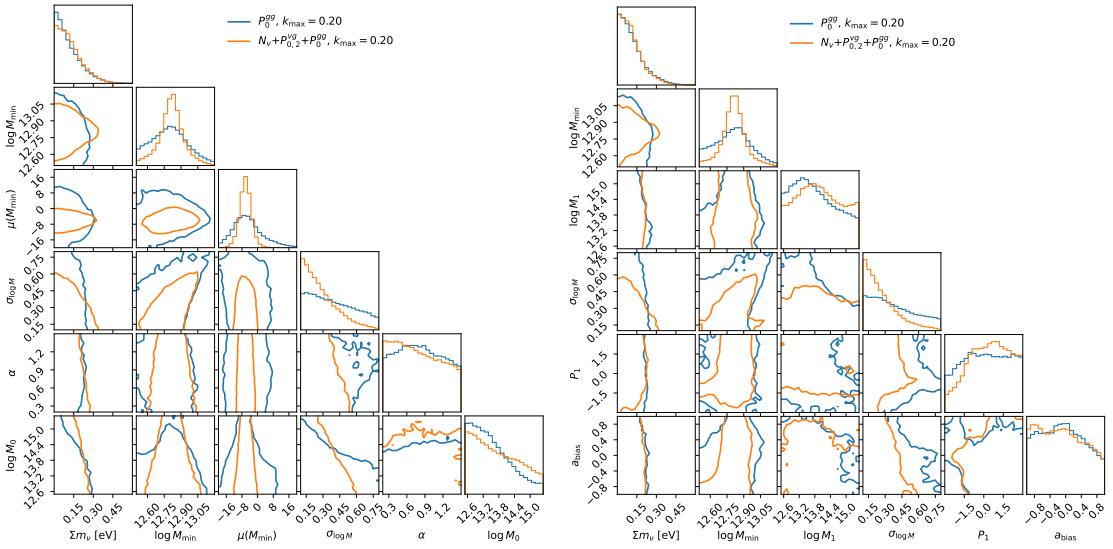


Figure F.5: Posteriors with different sets of nuisance parameters kept explicit. Since no hyperparameter optimization was performed when training the corresponding networks and the larger parameter space makes training more difficult, we assess these posteriors as less robust than our baseline results. The neutrino mass posteriors, however, are quite consistent. The left panel corresponds to the solid lines in Fig. F.1, the right panel to the dashed lines.

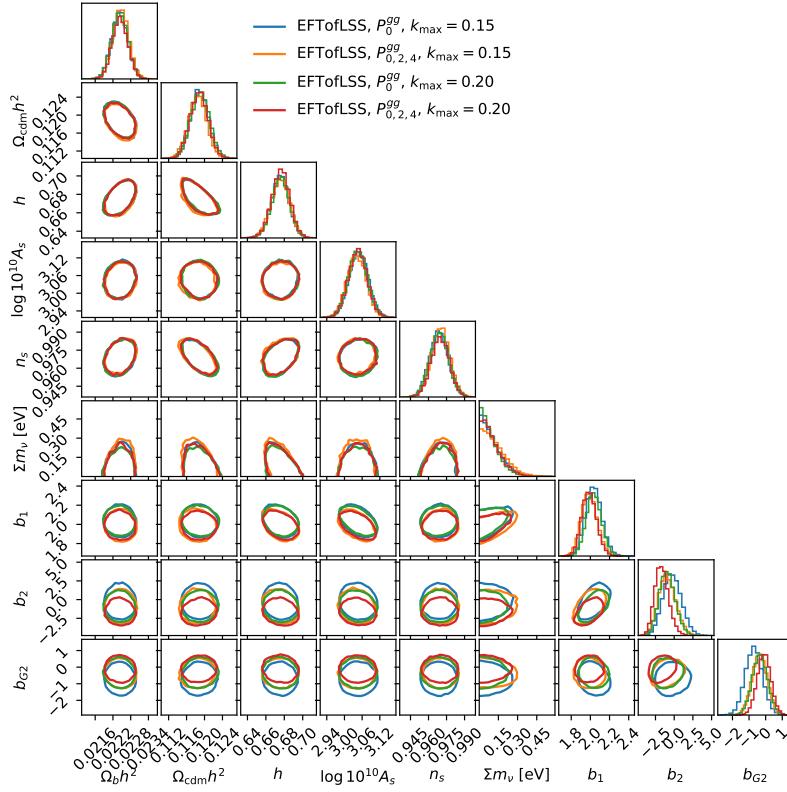


Figure F.6: For reference, we show the full EFTofLSS posteriors.

F.3 Corner plots

This appendix collects posteriors in the full parameter spaces considered. Fig. F.4 shows the baseline parameter space with two explicit nuisance parameters. Fig. F.5 shows larger sections of parameter space (it should be mentioned, however, that the corresponding neural networks were trained without further hyperparameter optimization, implying a somewhat lower level of confidence in the validity of these posteriors). Fig. F.6 shows our EFTofLSS posteriors, demonstrating that the Λ CDM part of the parameter space is prior-dominated.

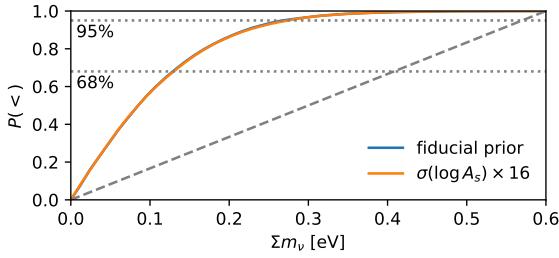


Figure F.7: Impact of relaxing the prior on A_s to an essentially uninformative level. The posterior on $\sum m_\nu$ remains essentially unchanged. This test is performed with the EFTofLSS likelihood since the implicit-likelihood setup would require significant computational expense for a post-hoc change of the prior.

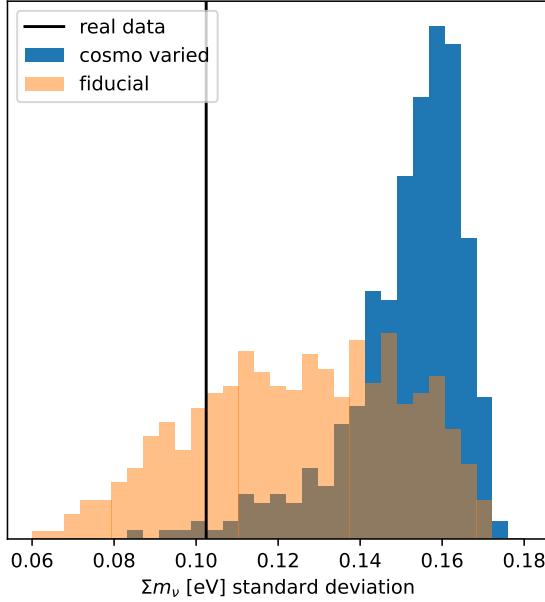


Figure F.8: Comparison of standard deviations of neutrino mass posteriors, run on cosmovaried and fiducial mocks.

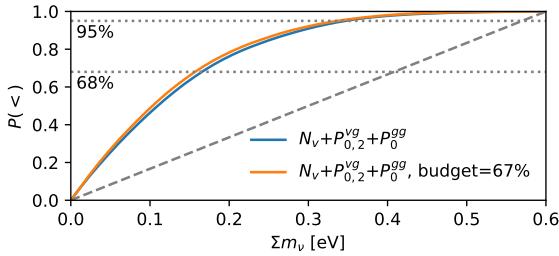


Figure F.9: Comparison of posteriors obtained from neural networks trained with the full training set and with 2/3 of the training set. We observe very good agreement between these posteriors.

F.4 Impact of A_s prior

F.5 Another look at the augmentations

F.6 Simulation budget

One might worry that the 127 cosmo-varied simulations are not enough to properly sample the cosmological prior. We test this by discarding a third of the simulations and training on the rest. The resulting posterior, compared to our baseline result, is shown in Fig. F.9. Agreement between the two posteriors is almost perfect, demonstrating that our simulations cover the cosmological prior sufficiently well.

F.7 Simulation data

About 50TB of halo catalogs, light cones, void catalogs, and summary statistics have been saved (at 20 times between $z = 0.44$ and $z = 0.68$ in 127 different massive-neutrino cosmologies with various HODs and 69 different initial conditions with a fiducial model). We are currently finalizing how to make this data set publicly available.

F.8 Code

In terms of new code, we have written C++ code to populate halo catalogs with galaxies and to generate light cones including survey realism.

We have also written a C implementation of the quasi-random sampling scheme for uniform and Gaussian priors.

This work necessitated several small modifications to public codes:

- `REPS`: read files generated by the current `CLASS` version; write output in a user-defined directory.
- `FastPM`: do not write neutrinos to disk.
- `bigfile`: support for half-precision floats.
- `Rockstar`: native reading of the `bigfile` snapshots generated by `FastPM` (using the file chunking to read in distributed fashion since `Rockstar` does not use MPI).
- `Rockstar/find_parents`: output to `bigfile` with lower priority fields in half precision.
- `cuboidremap`: support for velocities.
- `sbi`: custom splitting into training and validation data.

Since all these items are relatively obscure, we do not provide documentation. However, we are happy to share any of these with interested researchers. A repository with most of the code is available at https://github.com/leanderthiele/nuvoid_production.