
Forward modelling the large-scale structure
from the effective field theory
to
dark matter constraints and future survey optimization

Andrija Kostić



München 2023

Forward modelling the large-scale structure
from the effective field theory
to
dark matter constraints and future survey optimization

Andrija Kostić

Dissertation
der Fakultät für Physik
der Ludwig-Maximilians-Universität
München

vorgelegt von
Andrija Kostić
aus Leskovac, Serbien

München, den 31.5.2023

Erstgutachter: Prof. Dr. Eiichiro Komatsu
Zweitgutachter: Prof. Dr. Daniel Grün
Tag der mündlichen Prüfung: 17. Juli 2023

Contents	v
List of figures	viii
List of tables	ix
Zusammenfassung	xi
Abstract	xii
1 Introduction	1
2 Structure formation: Theoretical minimum	5
2.1 Standard model of structure formation	5
2.2 The WIMP dark matter	13
2.3 Dark matter as an effective fluid	16
2.4 Modelling biased tracers	25
2.5 Statistical inference	32
2.5.1 Hamiltonian Monte Carlo	35
2.5.2 Slice sampling	37
2.6 The field-level approach to structure formation	38
2.6.1 BORG forward model	39
2.6.2 EFTofLSS field-level forward model	41
3 Field-level inference with the EFT likelihood	45
3.1 Introduction	46
3.2 Forward models	47

3.2.1	LINEAR forward models	48
3.2.2	1LPT and 2LPT forward models	49
3.2.3	Field-level likelihood	51
3.3	Sampling the full posterior	53
3.3.1	Full posterior	53
3.3.2	Sampling methods	55
3.3.3	Code implementation	57
3.4	Synthetic datasets	58
3.4.1	LINEAR model synthetic data	58
3.4.2	2LPT synthetic data	59
3.5	Consistency test results	60
3.5.1	LINEAR forward models	61
3.5.2	1LPT and 2LPT forward models	71
3.6	Conclusions and Summary	77
3.7	Appendix A: Fourier space convention	79
3.8	Appendix B: Gaussian expectation for \hat{s} posterior	80
3.8.1	Appendix B.1: Linear model	80
3.8.2	Appendix B.2: Quadratic model	82
3.9	Appendix C: Running of b_{δ^2} with the cutoff	85
3.10	Appendix D: Convergence and sample correlation	89
3.11	Appendix E: 1LPT and 2LPT 2nd order bias coefficients	91
3.12	Appendix F: Marginalized and unmarginalized likelihood	93
4	Constraining WIMP dark matter using the local universe	95
4.1	Introduction	96
4.1.1	Theoretical background	98
4.2	Methods	101
4.2.1	Bayesian large-scale structure inference	101
4.2.2	Smoothed density field contribution	101
4.2.3	Computing the s -wave J factor	103
4.2.4	Computing the p - and d -wave J factor	109
4.2.5	Gamma ray data	112
4.2.6	Likelihood model	113
4.3	Results	118
4.3.1	The s -wave constraints	118
4.3.2	The p -wave and d -wave constraints	123
4.4	Interpretation of the results	126
4.4.1	Annihilation	126
4.4.2	Decay	129
4.5	Systematic uncertainties	130
4.5.1	J and D factor calculation	130

4.5.2	Constrained simulation volume	134
4.5.3	Cluster masses	136
4.5.4	Astrophysical templates	137
4.5.5	Thermal relics	137
4.5.6	Comparison to the literature	139
4.6	Conclusions	143
4.7	Appendix A: Calculation of J factor for d -wave	144
5	Information gain maps from the local universe	147
5.1	Introduction	148
5.2	Results	150
5.3	Discussion	155
5.4	Appendix A: 2M++ galaxy catalog	158
5.5	Appendix B: Fisher information maps	158
5.6	Appendix C: Marginalizing phase realizations and targeted searches . . .	162
6	Conclusion	165
A	Moments of the cold dark matter Boltzmann equation	167
B	Lagrangian perturbation theory	171
	Acknowledgements	203

LIST OF FIGURES

2.1	Thermal history of a $m_\chi = 100$ GeV WIMP	14
3.1	LEFTfield flowchart	56
3.2	Corner plots for FREEIC and FIXEDIC chains from Eq. (3.3) forward model on $\mathcal{D}_1^{\text{LINEAR}}$ dataset	62
3.3	Trace plot of the ICs for forward model from Eq. (3.3)	64
3.4	Posterior pdf of the ICs for forward models from Eqs. (3.3)–(3.5)	66
3.5	Corner plots for FREEIC and FIXEDIC chains for Eq. (3.5) forward model on $\mathcal{D}_2^{\text{LINEAR}}$	67
3.6	$b_{\delta^2, \Lambda}$ a posteriori mean running	69
3.7	Comparison of reconstructed ICs power spectrum running between Eqs. (3.3)–(3.5) forward models	70
3.8	Corner plot for FREEIC chains for 1LPT and 2LPT forward models on $\mathcal{D}^{2\text{LPT}}$ dataset	72
3.9	Comparison of auto-correlation functions for 2LPT forward models with different likelihood choices from Eqs. (3.15)–(3.16)	74
3.10	Posterior pdf of the ICs for 2LPT forward model from Eq. (3.11)	75
3.11	Corner plot for FREEIC chains for 2LPT forward model on $\mathcal{D}_2^{2\text{LPT}}$ dataset	76
3.12	Cornerplots for FREEIC chains for 1LPT and 2LPT forward models with different likelihood choices from Eqs. (3.15)–(3.16) on $\mathcal{D}_{1,b}^{2\text{LPT}}$ dataset	93
4.1	Mollweide projection of the s -wave annihilation and decay templates	103
4.2	The s -wave J factor and corresponding variance for NFW profile	107
4.3	Angular distribution of normalized J factor for a halo from CSiBORG suite	112
4.4	Posterior for 9844 CSiBORG realization for the s -wave channel	119
4.5	One-dimensional posterior on the J and D factor template amplitudes for the s -wave channel	119

4.6	One-dimensional posterior for the spectral amplitudes for s -wave channel	121
4.7	The s -wave DM annihilation cross section constraints	122
4.8	The DM decay rate constraints	123
4.9	The posterior distribution of spectral amplitudes for p - and d -wave channels	124
4.10	The p - and d -wave annihilation cross section constraints	125
4.11	The s -wave annihilation cross section constraints as a function of halo mass	127
4.12	The p - and d -wave annihilation cross section constraints as a function of halo mass	128
4.13	Thermal relic calculation for s -, p - and d -wave channels	138
5.1	BORG forward model schematic	149
5.2	Sensitivity maps of cosmological parameters	151
5.3	BORG posterior weighted Fisher information map for cosmological pa- rameters	152
5.4	Posterior weighted Fisher information map of the Coma cluster	154
5.5	Targeted searches flowchart	156

LIST OF TABLES

3.1	Parameters for the synthetic datasets	58
3.2	Gelman-Rubin statistics for chains part 1	87
3.3	Gelman-Rubin statistics for chains part 2	88
3.4	Gelman-Rubin statistics for chains part 3	88
4.1	Priors on dark matter properties	117

ZUSAMMENFASSUNG

Die Datenmenge künftiger Galaxiendurchmusterungen hat die Entwicklung immer besserer Techniken zur Gewinnung kosmologischer Informationen aus diesen Datensätzen vorangetrieben. Die Erhöhung der erwarteten Anzahl von Moden, die innerhalb der Reichweite der Theorie liegen könnten, bietet eine Verbesserung um mehrere Größenordnungen im Vergleich zum kosmischen Mikrowellenhintergrund (CMB). Diese zusätzliche Information ist in den nichtlinearen Strukturen verborgen. Um mit den anstehenden Datensätzen verantwortungsvoll umgehen zu können, ist es notwendig, die verschiedenen physikalischen Faktoren, die dabei eine Rolle spielen, sehr sorgfältig zu modellieren. Das Hauptziel dieser Arbeit war es daher, die Entwicklung und das Verständnis solcher theoretischen Modelle für die kosmische Strukturbildung voranzutreiben.

Kapitel 3 der Dissertation widmet sich dem Ansatz der effektiven Feldtheorie für die Strukturbildung, der unter dem Namen EFTofLSS bekannt ist. Der Ansatz wird auf der Feldebene angewandt, d.h. es wird versucht, maximale Informationen aus allen verfügbaren Moden in den Daten zu extrahieren. Es wurde eine neuartige Methode zur Durchführung von Selbstkonsistenztests des EFT-Vorwärtsmodells entwickelt, um sowohl die Reichweite des EFTofLSS-Ansatzes zu demonstrieren als auch die Leistung des Modells in verschiedenen Szenarien mit unterschiedlichen Graden von Modellfehlspezifikationen zu testen. Alle Modelle wurden auf eine Reihe von synthetischen Datensätzen angewandt. Daher sollte die in diesem Teil vorgestellte Arbeit als ein weiterer Schritt in Richtung einer kosmologischen Analyse auf voller Feldebene unter Verwendung des EFTofLSS-Ansatzes betrachtet werden.

Der nächste Teil, Kapitel 4, konzentriert sich auf die Demonstration des Potenzials von Rekonstruktionen der lokalen großräumigen Struktur auf Feldebene zur Einschränkung der fundamentalen Physik. Die Arbeit konzentriert sich auf die Extraktion von Grenzwerten für die Annihilations- und Zerfallsquerschnitte der dunklen Materie unter Verwendung der aus den beobachteten Galaxienkatalogdaten abgeleiteten Verteilung der dunklen Materie im lokalen Universum. Dies wird durch die Konstruktion von Schablonen (Templates) für die Annihilation und den Zerfall dunkler Materie aus dieser abgeleiteten Materieverteilung und die Kreuzkorrelation mit verfügbaren γ -Röntgendaten erreicht.

Kapitel 5 schließlich ist der Entwicklung von Ideen gewidmet, wo und wonach in zukünftigen Durchmusterungen gesucht werden sollte, um kosmologische Modelle effizienter einzuschränken. Es werden Karten vorgestellt, die Himmelsregionen nach ihrem posterior-gewichteten Informationsgewinnpotenzial für die jeweilige Gruppe von Modellen klassifizieren. Die Arbeit zeigt, dass der Inferenz auf Feldebene auf diese Weise einen Einblick gibt, wie zukünftige Suchstrategien aussehen sollten und wie sie schrittweise mit neuen kosmologischen Datensätzen im Rahmen von “active learning” verfeinert werden können.

The constraining power promised by future large-scale structure LSS surveys has driven the development of ever better techniques for extracting cosmological information from those datasets. Increase in the expected number of modes that could be well within the reach of the theory offers an improvement of few orders of magnitude with respect to cosmic microwave background (CMB). This extra information is hidden within the non-linear structures of the LSS. It is necessary to very carefully model different physics at play in order to responsibly deal with the upcoming datasets. Consequently, the main goal of this thesis was to push the development and understanding of such theoretical models for the clustering of the large-scale structure.

The Chapter 3 of the thesis is dedicated towards bringing closer LSS models utilizing the effective field theory approach, known collectively as EFTofLSS, to realistic galaxy-survey data applications. The approach is applied at the field-level, meaning that it tries to extract maximal information from all available modes in the data. A novel methodology for performing self-consistency tests of the EFT forward model framework was developed, both for demonstrating the reach of EFTofLSS approach, as well as test model performance in different scenarios with varying degrees of model mis-specification. All the models were applied to a set of synthetic datasets and hence the work presented in this part should be viewed as yet another step towards a final full field level cosmological analysis utilizing the EFTofLSS approach.

Next part, Chapter 4, is focused on demonstrating the potential of the field level reconstructions of the local large-scale structure in constraining fundamental physics. The work focuses on extracting the dark matter annihilation and decay cross-section constraints by using the inferred distribution of the dark matter in the local universe from the observed galaxy catalogue data. This is achieved by constructing templates of dark matter annihilation and decay from this inferred matter distribution and cross-correlating with available γ -ray data.

Finally, Chapter 5, is dedicated to developing ideas for where and what to look for in the future surveys in order to more efficiently constrain particular forward models of the LSS. Maps classifying regions of the sky according to their posterior weighted information gain potential are presented for the particular set of forward models. The work demonstrates that the field level approach gives, as a side product, an insight into how future search strategies should look like and how they can progressively be refined with new cosmological data sets within an active learning framework.

It is fair to say that the modern cosmology is built on top of the discovery of the Cosmic Microwave Background. From the initial discoveries of the Cosmic Microwave Background (CMB) anisotropies by the COBE satellite [1], over the subsequent introduction of the field of cosmology into the realm of precision science by the WMAP project [2, 3], up to pushing the concordance Λ CDM model predictions to ever higher multipoles on the sky by the Planck satellite [4, 5], the CMB continues to be the driving force of the bulk of cosmological research. However, the promise of an information gain of up to few orders of magnitude, offered by the future large-scale structure surveys such as DESI [6], Euclid [7], PFS [8], SPHEREx [9] and the LSST survey [10], to name a few, offers hope for a completely new regime for testing current cosmological theories. For example, taking $z \sim 1.6$, number of modes the large-scale structure offers is around $N_{\text{LSS}} \sim (k_{\text{max}}/k_{\text{min}})^3 \sim 10^9$, for wavelengths up to $k_{\text{max}} \sim 0.3 h \text{ Mpc}^{-1}$, while for CMB, taking $N_{\text{CMB}} \sim l_{\text{max}}^2 \sim 10^6$. This extra information however is hidden within the non-linear structures of our universe, such as the filaments, dark matter halos and galaxy clusters. This calls for careful modelling in order to be able to extract useful physics out of these future datasets.

The traditional approach to extracting cosmological information from large-scale structure surveys relies upon the analysis of summary statistics of the underlying galaxy density field. Mainly, focus is on the 2-, 3- and 4-point statistics (see [11–18]). In contrast to the angular power spectrum of the CMB, the main observable here is the 3D galaxy power spectrum. Some of the difficulties that affect these approaches are the unbiased estimation of the power spectrum from a set of observed galaxies on the sky and removal of survey mask effects which are themselves localized in real space, but become quite difficult to handle for higher order n -point correlation function estimators in Fourier space (e.g. [13]).

Alternative, advocated in this thesis, aims to extract information from all available Fourier modes offered by the large-scale structure surveys. This forward modelling technique allows for incorporating any physical effect, be it observational systematic or a new phenomenological piece. They are collectively known as *field-level forward modelling* approaches. Given that it is unclear how much cosmological information there is beyond the summary statistics approaches, it is a worthwhile effort to construct these models, and explicitly quantify the information gain.

One such example is the physical Bayesian forward modelling approach (BORG) [19–23]. This model is capable of producing realizations of present day galaxy count distribution on the sky by explicitly marginalizing over plausible initial conditions consistent with the observed galaxy data. As a side product, the model produces an ensemble of possible dark matter distributions in the local universe, which can be used to constrain fundamental dark matter properties. This we demonstrate in Chapter 4, motivating the use of such forward models beyond their initial goal of constraining cosmological models of structure formation. However, the cost of this modelling approach, among other things, is its very high dimensionality. In order to mend this problem, specific statistical inference techniques need to be used. We give a brief overview of such techniques in Section 2.5. Alongside this, it is not clear whether the mapping between the underlying dark matter density field and the observed galaxy tracer field is appropriate for all modes of interest. Given that we do not observe dark matter halos on the sky directly, but rather galaxies, it is necessary to have a well tuned model of the mapping between the underlying matter field and galaxy density field in order to avoid biasing the cosmological parameters of the forward model. Typically, what is assumed in BORG is a semi-empirical model from [24], which, as we explain in Section 2.4 does not give a complete picture. It is also not clear whether the model from Neyrinck et al. (*ibid.*) is well suited for all the Fourier modes BORG forward model aims to extract cosmological information from. Nevertheless, attempts have been made towards making this mapping more flexible, utilizing neural network approaches, albeit with limiting success [25, 26].

Another example of a structure formation forward modelling approach is the effective field theory of large-scale structure (EFTofLSS) [27–29]. This approach offers a rigorous and systematic framework, which allows for incorporating the physics of galaxy formation all the way up to quasilinear scales ($k_{\text{NL}} \sim 0.3 h \text{ Mpc}^{-1}$ at $z = 0$). In contrast to the BORG modelling framework, which utilizes a differentiable N-body simulation as its dark matter structure formation model, and a semi-empirical matter-tracer mapping, the EFTofLSS approach provides a perturbative model. The model is capable of producing the matter density field at redshift zero to a sub-percent accuracy when compared to N-body simulation result up to k_{NL} (see [30]), while as well offering matter-tracer mapping up to a given order in matter perturbations for any tracer of interest. This comes as a natural consequence of the equivalence principle, as we outline in Section 2.4 (see also [31]). The perturbative approach not only allows for robust structure formation model

all the way up to quasilinear scales, but also motivates the form the model likelihood, by explicitly marginalizing out the small scale physics of the structure formation problem. As shown in Section 2.6.2, this likelihood can be seen as an extension of the likelihood assumed in the **BORG** forward model (presented in Section 2.6.1). In Chapter 3 we make an effort towards applying this EFTofLSS approach at the field-level, and demonstrate its robustness in scenarios with model-mismatch between the generated synthetic data and the forward model employed. We develop a novel code capable of jointly sampling the cosmological parameters, EFT parameters of the model and marginalizes over the plausible initial conditions realizations at the same time. The following Chapter 2 serves as a broad overview of the prior knowledge required to understand the EFTofLSS modelling efforts that were undertaken in this thesis.

After providing a robust forward model and obtaining wanted constraints on cosmology, a natural question is where to look next for more data in order to further constrain the models at hand. The usual approach undertaken by most of the aforementioned surveys is to try to scan as big part of the sky as possible and within as large of a volume as possible. This takes a substantial effort both in terms of time and resources. In Chapter 5 we propose an idea for an automated active learning framework, which can offer a more efficient path towards constraining a favorite model at hand. It is not a novel concept, in the past similar ideas have been applied to CMB data [32], as well as survey geometry design [33]. However, the approach we propose in Chapter 5 is to also use the previously obtained model posterior, using it as a weighting measure for quantifying relative importance of the regions in the sky in terms of their information gain on our model parameters.

Additional work

During the course of the thesis, I have also collaborated on developing novel Causal inference models within the *Information Field Theory* framework [34, 35]. This was work built on top of my MSc thesis research and I have contributed to further development of the NIFTy code [36] by incorporating important pieces of the generative causal model structure and Bayesian model selection calculations. We applied this generative model to inferring causal relationship between patient’s viral load of the COVID-19 virus and patient’s age in [37] and extended the model by going beyond the additive noise assumption in [38]. Even though phenomenologically these papers are completely unrelated to the topic covered in this thesis, the methodology underlying the generative model framework itself has significant overlap.

CHAPTER 2

STRUCTURE FORMATION: THEORETICAL MINIMUM

In this section, we outline the theoretical framework behind structure formation modelling. We present the concordance structure formation model (Section 2.1) and its corrections within the framework of effective field theory of large-scale structure (Section 2.3). In Section 2.2 we use the derived phase-space dynamics of dark matter particles to study possible beyond Standard Model candidates for dark matter using the large-scale structure. This further motivates our work presented in Chapter 4 on constraining fundamental properties of dark matter. Then, we pursue towards a formulation of the theory of biased tracers (Section 2.4), heavily relied upon in the rest of the chapters of this thesis. Finally, we give a short overview of the necessary ingredients of the reasoning framework employed throughout and conclude the chapter by presenting two contrasting forward modelling frameworks in Section 2.6, explaining their advantages over the usually employed summary statistics approaches.

2.1 Standard model of structure formation

The standard cosmological model is built upon understanding the consequences of the theory of general relativity, fixing the initial conditions to the ones suggested by quantum field theory. The seminal work of V. Mukhanov and G. Chibisov [39] demonstrated that with the help of the inflationary phase in the early universe, initial quantum fluctuations can grow into cosmological perturbations and cause the subsequent formation of dark matter halos and galaxy clusters (see for example Chapter 8.3 of [40]). This even further consolidated the motivation for the inflation paradigm, besides it already providing the solutions to the well known horizon and flatness problems (see [40–42]). In fact, these initial fluctuations are also responsible for the anisotropies seen first by the COBE satellite [1]. However, the perturbations are rather small (of the order of $\sim 10^{-5} - 10^{-4}$) and

very closely respect statistical homogeneity and isotropy. This motivates that presently, on large-scales, the universe is homogeneous and isotropic as well, which defines the *cosmological principle*. Guided by the cosmological principle, it is then possible to uniquely fix the form of the metric tensor up to a coordinate transformation to (see Chapter 13.5 in [43])

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

with

$$\gamma_{ij}dx^i dx^j = a^2(t) \left(\frac{dr^2}{1 - kr^2} + r^2 d\Omega^2 \right), \quad (2.2)$$

where the $k \in \{-1, 0, +1\}$ represents the curvature constant, which has to be determined observationally, $a(t)$ represents the scale factor, r the radial coordinate distance¹ and $d\Omega^2 = d\theta^2 + \sin^2(\theta)d\phi^2$. Given the strong preference for a flat universe provided by the Planck team [5], from now on we will fix $k = 0$. The scale factor $a(t)$ can only be a function of time in order to preserve the homogeneity and isotropy of the spatial slices (being a maximally symmetric subspace). This is the well known *Friedmann-Robertson-Walker* (FRW) metric [44–50].

Knowing the metric not only allows us to write down the left-hand side of the Einstein field equation, but also fixes the form of the energy-momentum tensor. The field equations are given by

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = 8\pi GT_{\mu\nu}, \quad (2.3)$$

with $R_{\mu\nu}$ being the Ricci tensor, R the Ricci scalar, $T_{\mu\nu}$ the energy-momentum tensor and G being the gravitational constant. The expressions for the Ricci tensor and the Ricci scalar can be found in [41, 43]. Now, since we know the metric has to be form invariant, so must be any quantity that is derived from it. This implies then that the Ricci tensor and the Ricci scalar must be form invariant as well (see Chapter 13.4 in [43]). Furthermore, from the Equivalence Principle, it follows that any tensor involved in a covariant equation with a form invariant object has to be form invariant. Therefore, given the form of the field equations above, it follows that the energy-momentum tensor is also form invariant itself. The form of the FRW metric then fixes the energy-momentum tensor to be given as ²

$$T_{\mu\nu} = (\rho + p)u_\mu u_\nu + g_{\mu\nu}p, \quad (2.4)$$

with $\rho \equiv \rho(t)$ and $p \equiv p(t)$ being the background density and pressure respectively. Note that the background density, ρ , and pressure, p , can be only a function of time since they need to transform as scalars with respect to the symmetries of the spatial slices (rotations and translations). It should be noted that in presence of perturbations

¹Note that in this formulation all units are absorbed by the scale factor $a(t)$, while r is the coordinate on a grid scaled by $a(t)$.

²See for example discussion around Eq. 1.1.31 and Appendix B in [41]

around this background, the energy-momentum tensor will have additional terms such as anisotropic stress tensor or bulk viscosity (see for example [27, 51]). We will see the implications of this for cold dark matter evolution in Section 2.3.

Given this and conservation of the energy-momentum tensor, one can relate the scale factor a to the density and pressure of the background. The conservation equation implies

$$T^{\mu\nu}{}_{;\nu} = T^{0\nu}{}_{;\nu} + T^{i\nu}{}_{;\nu} = T^{0\nu}{}_{;\nu}, \quad (2.5)$$

since the covariant derivative of the second term in the first equality, $T^{i\nu}{}_{;\nu}$ is zero. To see this, one can expand the term as

$$T^{i\nu}{}_{;\nu} = T^{i0}{}_{;0} + T^{ij}{}_{;j} = T^{ij}{}_{;j} = g^{ij}{}_{;j} p(t). \quad (2.6)$$

The $T^{i0}{}_{;0}$ term is zero in the comoving fluid frame, since there $u^\mu = (1, \vec{0})$ and hence $T^{i0} = 0$, hence its covariant derivative is too. Given that this is a covariant equation, it holds in all frames. The term $g^{ij}{}_{;j}$ is the covariant derivative of the metric, i.e.

$$g^{ij}{}_{;j} = \partial_j g^{ij} + \Gamma_{jk}^i g^{kj} + \Gamma_{kj}^j g^{ki}, \quad (2.7)$$

with Γ being the Christoffel symbol

$$\Gamma_{\nu\alpha}^\mu = \frac{1}{2} g^{\mu\lambda} (\partial_\alpha g_{\lambda\nu} + \partial_\nu g_{\lambda\alpha} - \partial_\lambda g_{\alpha\nu}) \quad (2.8)$$

In order to see the covariant derivative of the metric is indeed zero, one can notice that in a local frame the g^{ij} corresponds to the flat-space Minkowski metric, for which the equality trivially holds. Therefore, given that it is a covariant equation and holds in one frame, it holds in all frames. Hence, the Eq. (2.5) simplifies to

$$\begin{aligned} T^{\mu\nu}{}_{;\nu} &\equiv T^{0\nu}{}_{;\nu} = \partial_t T^{00} + \Gamma_{\nu\mu}^0 T^{\mu\nu} + \Gamma_{\nu\mu}^\nu T^{\mu 0} \\ &= \frac{\partial \rho(t)}{\partial t} + 3 \frac{\dot{a}}{a} (p(t) + \rho(t)) \stackrel{!}{=} 0, \end{aligned} \quad (2.9)$$

from which it follows that $\rho \sim a^{-3(1+\kappa)}$ after assuming the equation of state of the form $p = \kappa\rho$. For different components of the cosmological energy budget, κ takes different values. For example, for the radiation $\kappa = 1/3$, for pressureless matter $\kappa = 0$, while the cosmological constant corresponds to an equation of state with $\kappa = -1$.

The Eq. (2.9) can be supplemented by the field equations from Eq. (2.3), taking the expression of the energy momentum tensor from Eq. (2.4). This then gives (see for example [41]) the *Friedmann equations*

$$\begin{aligned} \left(\frac{\dot{a}}{a}\right)^2 &= \frac{8\pi G}{3} \rho \\ \frac{\ddot{a}}{a} &= -\frac{4\pi G}{3} (3p + \rho). \end{aligned} \quad (2.10)$$

These equations govern how the background cosmology evolves with time. One can immediately notice from the first equation that the scale factor, a , will continuously increase with time if $\rho > 0$. In fact, the first of the two equations from Eq. (2.10) is usually written as

$$H^2 = H_0^2(\Omega_r a^{-4} + \Omega_m a^{-3} + \Omega_\Lambda) \quad (2.11)$$

with

$$\Omega_i = \frac{8\pi G}{3H_0^2} \rho_{0,i} \quad (2.12)$$

denoting the fraction of the energy budget stored within different components of the universe, with $\rho_{0,i}$ being the present density of different components. For example, matter ($i \equiv m$), radiation ($i \equiv r$) or cosmological constant contribution ($i \equiv \Lambda$). We also implicitly assumed the present day scale factor is fixed to $a_0 = 1$. The $H \equiv \dot{a}/a$ is the Hubble parameter, while H_0^2 corresponds to its current value.

As already suggested, in order to extract as much of the physical information as possible from the large-scale structure, one needs to understand how to model the perturbations around this background cosmology, since it is clear that the universe doesn't look completely isotropic and homogeneous on all scales. This requires a careful treatment of all possible perturbations to the Einstein field equation Eq. (2.3). The usual approach is to decompose the perturbations into scalars, divergenceless vectors and divergenceless traceless symmetric tensors and write down to leading order how the perturbed field equations look in this case (see for example Chapter 5.1 in [41] or Chapter 7 in [40]). Since in the bulk of this thesis, the focus will be on the dark matter component of our universe, we will focus on the relevant physics governing its evolution in what follows. This means that it is necessary to study only the scalar perturbations of the metric.

In the conformal Newtonian gauge³, the scalar perturbations change the background metric as (for example Eq. 3.49 in [42])⁴

$$g_{00} = -1 - 2\Psi, \quad g_{i0} = 0, \quad g_{ij} = a^2 \delta_{ij} (1 + 2\Phi). \quad (2.13)$$

The scalar fields Φ and Ψ are not independent in case one can neglect the anisotropic part of the energy-momentum tensor, which is the case for example for the cold dark matter particles due to their small velocities on the scales of interest, as demonstrated in Section 2.3 (see Eq. (2.62) and discussion around it). In fact, neglecting the anisotropic part of the energy-momentum tensor allows for equating $\Phi = -\Psi$, and the field Ψ can be interpreted as the Newtonian potential arising from the matter density perturbations.

Given that we are not interested in tracking down each individual dark matter particle in the universe, it is useful to instead study the change in the corresponding phase-space density, or more feasibly, the moments thereof. The equation governing the evolution of

³For the consistent treatment in both synchronous and conformal Newtonian gauge see [52]

⁴Note that different conventions are used in [41] and [40] for example

the phase-space distribution function in time is the *Boltzmann equation*. Neglecting the self-interaction processes of the dark matter particles, due to a very small cross sections (very long decay rates), as demonstrated in Section 2.2 and Chapter 4, the relevant Boltzmann equation is of the collisionless type

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \frac{\partial f}{\partial x^i} \frac{dx^i}{dt} + \frac{\partial f}{\partial p^i} \frac{dp^i}{dt} \stackrel{!}{=} 0. \quad (2.14)$$

Furthermore, assuming the dark matter particles to be non-relativistic⁵, the geodesic equations for x^i and p^i simplify significantly, and the Boltzmann equation reduces to (see Chapter 12 in [42])

$$\frac{\partial f}{\partial t} + \frac{\partial f}{\partial x^i} \frac{p^i}{ma} - \frac{\partial f}{\partial p^i} \left(H p^i + \frac{m}{a} \frac{\partial \Psi}{\partial x^i} \right) = 0, \quad (2.15)$$

with m being the non-relativistic dark matter particle mass. Taking the first two moments of the above Eq. (2.15) (see also Appendix A) produces the following two equations

$$\begin{aligned} \frac{\partial}{\partial \eta} \rho + \frac{\partial}{\partial x^i} (\rho v^i) + 3\mathcal{H}\rho &= 0, \\ \frac{\partial}{\partial \eta} v^i + v^j \frac{\partial}{\partial x^j} v^i + \mathcal{H}v^i + \frac{\partial}{\partial x^i} \Psi + \frac{1}{\rho} \frac{\partial}{\partial x^j} \tau^{ij} &= 0 \end{aligned} \quad (2.16)$$

where the derivatives with respect to the proper time, t , have been switched into derivatives with respect to the conformal time, η defined through $dt = a(\eta)d\eta$ and $\mathcal{H} = aH$ being the conformal Hubble parameter. As can immediately be noticed, these equations are the continuity and Euler equations in the presence of the Hubble flow. What is instructive and usually done at this point (for example Chapter 12.2 in [42]) is to neglect the presence of the stress tensor τ^{ij} and continue solving the system treating the dark matter as a pressureless fluid, which happens to be quite a good description for cold dark matter (see Eq. (A.4) and discussion below it). However, dark matter is not a perfectly pressureless fluid, but represents a collection of collisionless particles and hence a more correct treatment is that of an *effective fluid* as discussed in Section 2.3. Nonetheless, solving the system of equations from 2.16 without the presence of τ^{ij} is still valid at the tree-level and hence coincides with the effective description at this order. Continuing along these lines, using the fact the background density evolves as given in Eq. (2.9) and supplementing the Poisson equation one obtains the following system of equations (see

⁵For example taking WIMPs (see Section 2.2) as candidates, with a typical mass close to $m_\chi \sim 100$ GeV, which is much higher than the temperatures throughout the history of the observed universe $\sim 10^{-4} - 10^{-1}$ eV

Eq. 12.23 in [42])

$$\begin{aligned} \delta' + \frac{\partial}{\partial x^i} ((1 + \delta)v^i) &= 0, \\ v^{i'} + v^j \frac{\partial}{\partial x^j} v^i + \mathcal{H}v^i + \frac{\partial}{\partial x^i} \Psi &= 0, \\ \nabla^2 \Psi &= \frac{3}{2} \mathcal{H}^2 \Omega_m(\eta) \delta, \end{aligned} \quad (2.17)$$

with $\Omega_m(\eta)$ being the time dependent matter density parameter, absorbing the time dependence of the background matter density, unlike the present time density parameters defined in Eq. (2.12). The primes denote derivatives with respect to the conformal time η .

It is useful to decompose the Euler equation into the divergence and curl part. The curl component produces the following equation ⁶

$$\partial_\eta \omega^i + \mathcal{H} \omega^i + \partial_j (v^j \omega^i) - \partial_j (v^i \omega^j) = 0, \quad (2.18)$$

using the identity $v^j \partial_j v^i = \epsilon^{ijk} v_j \omega_k + \frac{1}{2} \partial_i (v^j v_j)$ and defining $\omega^i = \epsilon^{ijk} \partial_j v_k$, with ϵ^{ijk} being the Levi-Civita pseudo-tensor. As can be seen, if the initial curl of the velocity vanishes, it remains zero. Given that vector perturbations of the metric decay both inside and outside the horizon it is safe to assume that the initial curl is very small ⁷. It is also good to note that even with the inclusion of the stress-tensor term, $\partial\tau/\rho$, into the Euler equation of Eq. (2.16) the curl will be sourced only at the second order in density perturbations, i.e.

$$\partial \times \left(\frac{1}{\rho} \partial\tau \right) = \partial \times \left(\frac{1}{\bar{\rho}(1 + \delta)} \partial\tau \right) \sim \partial \times (\partial\tau(1 - \delta)) \sim \partial \times (-\partial\tau\delta) \sim \delta^2$$

given that the stress-tensor itself is of order $\sim \delta$ (see Eq. (2.50) and discussion around it). Therefore, indeed at leading order the vorticity is sourced at $\sim \delta^2$, and hence the vorticity starts at $\sim \delta^3$ order.

Since we will be expanding up to the 2nd order in the discussion below, we don't need to worry about the curl component of the velocity field and instead we focus on the divergence part. This allows to write the velocity field as $v^i = \frac{\partial_i}{\nabla^2} \theta$, θ being the divergence of the velocity field. Therefore, the coupled system of PDEs to solve is

$$\begin{aligned} \delta' + \theta &= -\partial_i (v^i \delta), \\ \theta' + \mathcal{H}\theta + \frac{3}{2} \Omega_m(\eta) \mathcal{H}^2 \delta &= -\partial_i (v^j \partial_j v^i). \end{aligned} \quad (2.19)$$

⁶See also Mehrdad Mirbabayi's comprehensive yet concise lecture notes available at the following coordinates <https://users.ictp.it/~mirbabayi>

⁷However, in the presence of primordial magnetic fields, non-zero curl can be generated, although the field strength is most likely very small. See for example [53].

To linear order, the equations become

$$\begin{aligned}\theta^{(1)} + \partial_\eta \delta^{(1)} &= 0, \\ \partial_\eta^2 \delta^{(1)} + \mathcal{H} \partial_\eta \delta^{(1)} - \frac{3}{2} \Omega_m(\eta) \mathcal{H}^2 \delta^{(1)} &= 0.\end{aligned}\tag{2.20}$$

We see that at this order the two equations decouple, and that the velocity divergence is directly sourced by the linear density perturbations. In what follows we restrict to the EdS universe ($\Omega_m = 1$) given that the period during which structures form is very much matter dominated and the overall error made when compared to the full Λ CDM solution is at the percent level (see for example [30]). Having this in mind and switching from the conformal time derivative into a derivative with respect to the scale factor $\partial_\eta = a' \partial_a$ produces the following *linear growth equation*

$$\mathcal{H}^2 \left(a^2 \partial_a^2 + \frac{3}{2} a \partial_a - \frac{3}{2} \right) \delta^{(1)} = 0.\tag{2.21}$$

Looking at this second order ODE in Fourier space, unveils that at linear order in density perturbations, all the modes decouple and evolve in the same manner. The two solutions for the time dependence are given by $D_+ = a$ and $D_- = a^{-3/2}$. For systematic treatment to higher orders, it is useful to write down the corresponding Green's function of Eq. (2.21), which in EdS becomes

$$G(a > a_i, a_i) = \frac{D_-(a)D_+(a_i)}{\zeta(a_i)W(a_i)} - \frac{D_+(a)D_-(a_i)}{\zeta(a_i)W(a_i)} = \frac{2}{5H_0^2} \left(\frac{a}{a_i} - \left(\frac{a_i}{a} \right)^{3/2} \right),\tag{2.22}$$

with a_i denoting the starting value of the scale factor, at which the initial conditions are specified, and $\zeta(a_i) = (\mathcal{H}(a_i))^2 a_i^2$. The $W(a_i)$ represents the Wronskian of the differential equation evaluated at a_i . A corresponding expression for a general Λ CDM cosmology can be found in [54].

In the following we keep only the D_+ part, since it corresponds to the growing solution, which is the one relevant for the structure formation. Therefore, we see that at leading order the solutions are

$$\begin{aligned}\delta^{(1)}(\mathbf{k}, a) &= a \delta_1(\mathbf{k}), \\ \theta^{(1)}(\mathbf{k}, a) &= -a \mathcal{H} \partial_a \delta^{(1)}(\mathbf{k}, a),\end{aligned}\tag{2.23}$$

where $\delta_1(\mathbf{k})$ denotes the initial conditions. As we will see, $\delta_1(\mathbf{k})$ is fully specified with the expression from Section 3.2.1 (second line of Eq. (3.4)) and can be used in that form to set the initial conditions for our forward models.

To continue to higher orders, we can turn back to the full PDE system in Eq. (2.19), remember that we're keeping $\Omega_m = 1$, and rewrite it as

$$\begin{aligned}\theta &= -a \mathcal{H} \partial_a \delta - \partial_i(v^i \delta), \\ \mathcal{H}^2 \left(a^2 \partial_a^2 \delta + \frac{3}{2} a \partial_a \delta - \frac{3}{2} \delta \right) &= \partial_i(v^j \partial_j v^i) - \mathcal{H}(a \partial_a + 1) \partial_i(v^i \delta).\end{aligned}\tag{2.24}$$

Since we know the Green's function of the second PDE, we can straightforwardly obtain higher order solutions using perturbation theory. It is only necessary to keep track of all relevant source terms on the right hand side of the PDE at the given order. At second order, the solution for the divergence and density perturbation is

$$\begin{aligned}\delta^{(2)}(\mathbf{x}, a) &= a^2 \left(\frac{5}{7} \delta_1^2(\mathbf{x}) + \frac{2}{7} \left(\frac{\partial_i \partial_j}{\nabla^2} \delta_1(\mathbf{x}) \right)^2 + \partial_i \delta_1(\mathbf{x}) \left(\frac{\partial_i}{\nabla^2} \delta_1(\mathbf{x}) \right) \right), \\ \theta^{(2)}(\mathbf{x}, a) &= -a^2 \mathcal{H} \left(\frac{3}{7} \delta_1^2(\mathbf{x}) + \frac{4}{7} \left(\frac{\partial_i \partial_j}{\nabla^2} \delta_1(\mathbf{x}) \right)^2 + \partial_i \delta_1(\mathbf{x}) \left(\frac{\partial_i}{\nabla^2} \delta_1(\mathbf{x}) \right) \right)\end{aligned}\quad (2.25)$$

or equivalently in Fourier space

$$\begin{aligned}\delta^{(2)}(\mathbf{k}, a) &= a^2 \int_{\mathbf{k}_1} \int_{\mathbf{k}_2} \delta_D^{(3)}(\mathbf{k} - \mathbf{k}_1 - \mathbf{k}_2) F_2(\mathbf{k}_1, \mathbf{k}_2) \delta_1(\mathbf{k}_1) \delta_1(\mathbf{k}_2), \\ \theta^{(2)}(\mathbf{k}, a) &= -a^2 \mathcal{H} \int_{\mathbf{k}_1} \int_{\mathbf{k}_2} \delta_D^{(3)}(\mathbf{k} - \mathbf{k}_1 - \mathbf{k}_2) G_2(\mathbf{k}_1, \mathbf{k}_2) \delta_1(\mathbf{k}_1) \delta_1(\mathbf{k}_2),\end{aligned}\quad (2.26)$$

where the Fourier convention of Section 3.7 being utilized, the $\delta_D^{(3)}$ being the three-dimensional Dirac delta function and the symmetrized kernels are

$$\begin{aligned}F_2(\mathbf{k}_1, \mathbf{k}_2) &= \frac{5}{7} + \frac{2}{7} \frac{(\mathbf{k}_1 \cdot \mathbf{k}_2)^2}{k_1^2 k_2^2} + \frac{1}{2} \frac{\mathbf{k}_1 \cdot \mathbf{k}_2}{k_1 k_2} \left(\frac{k_1}{k_2} + \frac{k_2}{k_1} \right), \\ G_2(\mathbf{k}_1, \mathbf{k}_2) &= \frac{3}{7} + \frac{4}{7} \frac{(\mathbf{k}_1 \cdot \mathbf{k}_2)^2}{k_1^2 k_2^2} + \frac{1}{2} \frac{\mathbf{k}_1 \cdot \mathbf{k}_2}{k_1 k_2} \left(\frac{k_1}{k_2} + \frac{k_2}{k_1} \right).\end{aligned}\quad (2.27)$$

The same procedure can be repeated to higher orders and it can be shown by induction that the solution will be given as

$$\delta(\mathbf{x}, a) = \sum_{n=1}^{\infty} a^n \delta_n(\mathbf{x}), \quad \theta(\mathbf{x}, a) = -\mathcal{H} \sum_{n=1}^{\infty} a^n \theta_n(\mathbf{x}),\quad (2.28)$$

with $v_n = \frac{\partial_i}{\nabla^2} \theta_n$ and δ_n related through the following recursion relation

$$\delta_n = \frac{1}{(n-1)(n+3/2)} \sum_{\substack{m_1+m_2=n \\ m_1>0, m_2>0}} (\partial_i (v_{m_1}^j \partial_j v_{m_2}^i) - (n+1/2) \partial_i (\delta_{m_1} v_{m_2}^i)),\quad (2.29)$$

which can be obtained by directly plugging in the general solution from Eq. (2.28) into the Euler equation from Eq. (2.24). It is worth emphasizing once more that the solutions derived above are valid for the EdS universe, while the solutions for the Λ CDM model have slightly different form [54, 55]. Collectively, the perturbative solutions presented in

this chapter are known as *Eulerian perturbation theory* solutions. As already mentioned, it relies on the smallness of density fluctuations δ . It is however possible to go beyond this approximation in the *Lagrangian perturbation theory* (LPT) approach, as described in Appendix B. The EFT forward models we employ, as presented in Section 2.6.2 and Chapter 3, follow the LPT description. The two frameworks are of course equivalent, but have different convergence properties with respect to the fully non-linear solution for the velocity and density field obtained from N-body simulations (see Appendix B for discussion on these points). In the following, we stick to the Eulerian perturbation theory formulation for clarity.

2.2 The WIMP dark matter

Even though significant advances towards a complete theoretical description of dynamical evolution of the dark matter fluid have been made over the past few decades, the dark matter's intrinsic nature still stays elusive to this day. Over the years, there have been many proposals for potential candidates, motivated by different extensions of the Standard Model. The wealth of proposed candidates is well captured by the fact that the possible mass range spans 90 orders of magnitude [42]. One of the most promising candidates through the past few decades was the *weakly interacting massive particle*, WIMP for short. The WIMP has been the favorite cold dark matter candidate, arising from the extension of the electroweak sector of the Standard Model [56–58], for the reasons that will be discussed in this section. Among the wealth of different particle types WIMPs could take on, the Majorana fermion stayed the most favorable candidate (see Section 12 in [59] for example). Therefore, before continuing with further investigation into the clustering of dark matter, we can first utilize the derived Boltzmann equation of the previous chapter to study the background evolution of WIMP dark matter particles. This will provide further context for the pursuit of dark matter constraints from the local universe as performed in Chapter 4.

As in Section 2.1, we are interested in the averaged behavior of the dark matter particles, hence we start our study again from the Boltzmann equation. Instead of focusing on the perturbations around the background density, here we will focus on the average abundance of dark matter particles as a function of time. Therefore, the Eq. (2.15) reduces to

$$\frac{\partial f}{\partial t} - \frac{\partial f}{\partial p^i} H p^j = \mathcal{I}[f], \quad (2.30)$$

where we have dropped $\partial_i f$ term due to homogeneity, the metric perturbation term $\partial_i \Psi$ due to our focus on the background density and we have added the source term $\mathcal{I}[f]$. The source term, otherwise known as collision term, is present due to the interactions the WIMP dark matter particles are involved in. Here we focus on annihilation, since decay is treated analogously. To understand how these interactions change the background

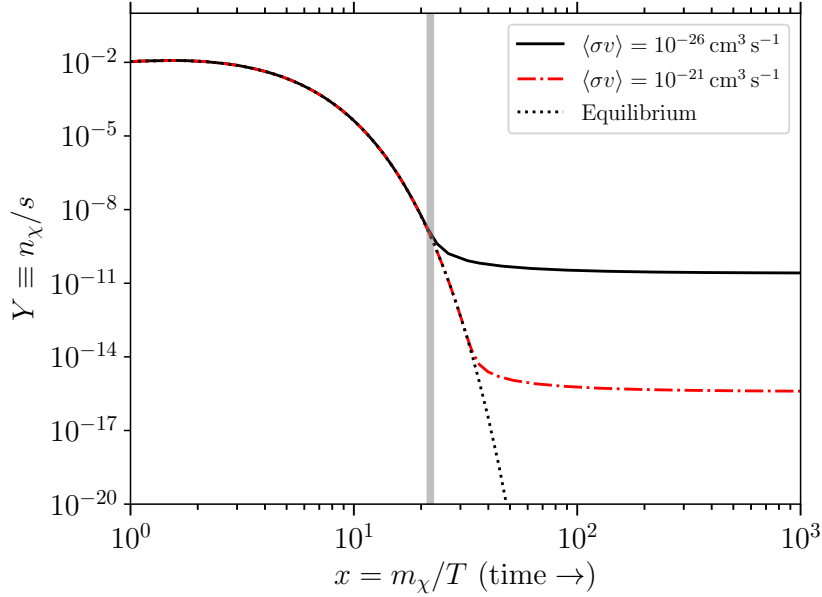


Figure 2.1: Thermal history of a dark matter WIMP, assuming it corresponds to a Majorana fermion, with a mass of $m_\chi = 100 \text{ GeV}$. The plot shows the change in WIMP abundance (n_χ/s) with time. The dotted line represents the evolution of the equilibrium number density, while the black and red represent the evolution of a WIMP thermal relic with different cross sections indicated in the top right of the figure. The gray line indicates the moment of freeze-out for the WIMP with cross section $\langle\sigma v\rangle = 10^{-26} \text{ cm}^3 \text{ s}^{-1}$.

abundance, we can integrate the phase-space distribution function f over momenta and arrive at the following equation

$$\frac{1}{a^3} \frac{d(n_\chi a^3)}{dt} = \int_{\mathbf{p}} \mathcal{I}[f], \quad (2.31)$$

where n_χ represents background dark matter particle density. The collision term can be derived from the following consideration. We focus on Majorana WIMPs, since they are the main candidate we consider in Chapter 4,. This implies the annihilation interactions will have the following form

$$\chi\chi \rightleftharpoons A\bar{A}, \quad (2.32)$$

where A and \bar{A} denote some standard model part-antiparticle pair and taking Majorana WIMPs are their own antiparticle, i.e. $n_\chi = n_{\bar{\chi}}$. It is then straightforward to derive the collision term purely from dimensional grounds. Given that the number of particle pairs $A\bar{A}$ generated from annihilation will be proportional to n_χ^2 and vice versa for the

opposite direction, Eq. (2.31) can be written as (see Chapter 4.1 in [42])

$$\frac{1}{a^3} \frac{d(n_\chi a^3)}{dt} = -\langle \sigma v \rangle n_\chi^2 + \alpha n_A n_{\bar{A}}. \quad (2.33)$$

The minus sign in front of the first term signifies that the budget of χ particles is depleted upon annihilation, while upon the reverse reaction it is enhanced and hence the plus sign. The $\langle \sigma v \rangle$ represents the thermally averaged annihilation cross section (see Eq. (4.77)). For s -wave annihilation, this is a straightforward calculation, while for p - and d -wave, the thermal cross section requires a bit more effort as described in Section 4.5.5. In the following, we focus only on s -wave annihilation, for which $\langle \sigma v \rangle$ is a constant velocity independent number. Next, we notice that in equilibrium we have

$$\alpha = \langle \sigma v \rangle \frac{(n_\chi^{\text{eq}})^2}{(n_A n_{\bar{A}})_{\text{eq}}}. \quad (2.34)$$

Assuming that the final products $A\bar{A}$ are much less massive than our WIMP particle χ , and essentially in complete equilibrium before WIMP goes out of equilibrium, we have $n_A = (n_A)_{\text{eq}}$ ($n_{\bar{A}} = (n_{\bar{A}}^{\text{eq}})$). This implies

$$\frac{1}{a^3} \frac{d(n_\chi a^3)}{dt} = -\langle \sigma v \rangle (n_\chi^2 - (n_\chi^{\text{eq}})^2), \quad (2.35)$$

which is the well known *Riccati* equation. Dividing out by the entropy density, s , removes the dependence on the scale factor, assuming adiabatic expansion. Furthermore, presuming most of the annihilation interactions occur during the radiation domination era we can recast from explicit time dependence to temperature dependence, since during radiation domination the two are related as

$$\left(\frac{T}{\text{MeV}} \right) \sim \left(\frac{t}{\text{s}} \right)^{1/2},$$

where temperature is measured in MeV and proper time in seconds. Taking $x \equiv m_\chi/T$ Eq. (2.35) becomes [60] (see also Section 4.5.5)

$$\frac{dY}{dx} = -\frac{s \langle \sigma v \rangle}{Hx} \left[1 + \frac{1}{3} \frac{d(\ln g_{\text{eff}}(T))}{d(\ln T)} \right] (Y^2 - Y_{\text{eq}}^2), \quad (2.36)$$

where $g_{\text{eff}}(T)$ is the number of effective degrees of freedom contributing to entropy density and has to be calculated phenomenologically from the thermodynamics of the Standard Model of particle physics. For example, reference [61] provides such a calculation (see their Fig. 4). Given that the range of temperatures we are interested in is $T \ll m_\chi$, the equilibrium comoving number density is given as the corresponding limit of the Fermi-Dirac distribution

$$Y_{\text{eq}} = \frac{45}{2\pi^4} \left(\frac{\pi}{8} \right)^{1/2} \frac{g_\chi}{g_{\text{eff}}} x^{3/2} \exp(-x), \quad (2.37)$$

where g_χ represents the number of internal degrees of freedom for the WIMP. For example, for the spin $1/2$ Majorana fermion WIMP, as usually assumed, this number is 2. With this, it is possible to numerically solve Eq. (2.35). The solution of this calculation is plotted in Fig. 2.1 for a WIMP of mass $m_\chi = 100$ GeV. The change in mass weakly affects the relic abundance, i.e. the comoving number density in the limit $m_\chi/T \rightarrow \infty$ (see for example Fig. 1 from [60]). As one can see from Fig. 2.1, the abundance stays almost unchanged after $x \sim 25$ (vertical gray band in Fig. 2.1), which corresponds to the freeze-out temperature for the WIMP of mass $m_\chi = 100$ GeV. It turns out that for $\langle\sigma v\rangle \sim 10^{-26} \text{ cm}^3 \text{ s}^{-1}$ (black line in Fig. 2.1) is able to roughly match the present day dark matter abundance of $Y \sim 10^{-9}$ [40]. Interpreting this result as coming from a particle annihilating with the cross section $\langle\sigma v\rangle \sim \alpha_\chi^2/m_\chi^2$ suggests a coupling constant of around $\alpha_\chi \sim 1/100$ which is close to the electroweak coupling constant $\sim 1/137$. This coincidence, among other favorable properties, was the reason why WIMP model became so popular in the first place. In Chapter 4 we use the inferred local large-scale structure dark matter distribution by the BORG forward model (see Section 2.6.1) in order to constrain this cross section for different annihilation channels.

2.3 Dark matter as an effective fluid

The approach presented in Section 2.1 is known as the *standard perturbation theory* (SPT). As mentioned previously, it is valid at tree-level only, as will be explained in this section. One can try nonetheless and apply the SPT approach to model the 2-point statistics of the dark matter halo density field, for example, extracted from N-body simulations. However, we don't observe dark matter halos but galaxies instead, and this requires an introduction of several more parameters to the model, collectively called *bias parameters* (see Section 2.4). Even so, using the 2-point statistics of the galaxy field can still provide useful information on cosmology, given the Equivalence Principle. In other words, the dark matter halos and galaxies cluster the same way, since they feel the gravitational field in the same way. It is therefore a worthwhile effort to try and see how well the SPT approach works in recovering the underlying cosmology given an N-body halo density field. This is precisely what was done in [62, 63]. The conclusion of this work was that the SPT is useful, but only up to relatively high redshift of $z \geq 1$. The reason being simply that the perturbation series from Eq. (2.28) doesn't converge anymore on a big enough range of scales in order for it to be a useful series expansion. The cause of such behavior is that the density perturbations become of order unity at ever larger scales, as time passes by and larger and larger structures form. This in turn reduces the applicability of the series expansion. In order to understand this in more detail, it is enough to look at the typical amplitude of the linear density fluctuations at

a given time t

$$\langle \delta^{(1)}(\mathbf{x}, t) \delta^{(1)}(0, t) \rangle = \int_{\mathbf{k}} P_L(k, t) e^{i\mathbf{k}\cdot\mathbf{x}}. \quad (2.38)$$

The quantity $P_L(k, t)$ is the linear power-spectrum at time t . It is straightforward to obtain this relation given the solution for $\delta^{(1)}$ from Eq. (2.23). The modulus of the integrand of the above equation is $\Delta(k) = a^2 k^3 P_L(k)$. Due to the shape of the linear power-spectrum, this is a monotonically increasing function of k . This means that one can find a typical scale $|\mathbf{x}| \sim R$ at which the integral becomes of order unity. This is precisely the scale up to which it will be valid to apply the perturbation series from Eq. (2.28), since for modes beyond the scale $k_{\text{NL}} \sim 1/R$, the SPT solution will start to deviate strongly from the correct non-linear solution for the halo clustering problem, which is given simply as a result of an N-body simulation. For redshifts up to $z \sim 1$ this is of order $k_{\text{NL}} \sim 0.24 h \text{ Mpc}^{-1}$ [62, 63], corresponding to a scale of about $R \sim 4 h^{-1} \text{ Mpc}$. During the subsequent clustering, this scale will only continue increasing as more and more modes will go into the non-linear regime, simply because larger and larger halos of dark matter will form.

One can then try to add higher n -point correlation functions into the fold, but as one can conclude from the above, their range of validity within the SPT approach will be constrained to an even smaller set of modes. Now, given that some of the forward modelling approaches do use a full N-body simulation as a part of the forward model, they will not have the same issues. They will however suffer from other problems, such as a very high number of degrees of freedom and the proper choice of the bias expansion, as will be explained in Section 2.6. This then raises a question whether it is possible to somehow regulate the perturbation series and extend its applicability to a higher range of scales in order for it to be useful at redshifts $z \sim 0$. This was demonstrated to indeed be possible, for the first time in [27–29], by developing the *effective field theory of large-scale structure* (EFTofLSS).

In order to grasp the main concepts of the EFTofLSS approach, it is useful to write out the 2-point function at next-to-leading order (again assuming EdS)

$$\begin{aligned} \langle \delta(\mathbf{k}, t) \delta(-\mathbf{k}, t) \rangle' &= a^2 \langle \delta_1(\mathbf{k}) \delta_1(-\mathbf{k}) \rangle' + a^4 (\langle \delta_2(\mathbf{k}) \delta_2(-\mathbf{k}) \rangle' + 2 \langle \delta_1(\mathbf{k}) \delta_3(-\mathbf{k}) \rangle') \\ &= a^2 P_L(k) + a^4 (P_{22}(k) + 2P_{13}(k)), \end{aligned} \quad (2.39)$$

where

$$\begin{aligned} P_{22}(k) &= 2 \int_{\mathbf{q}} F_2^2(\mathbf{q}, \mathbf{k} - \mathbf{q}) P_L(q) P_L(|\mathbf{k} - \mathbf{q}|), \\ P_{13}(k) &= 3P_L(k) \int_{\mathbf{q}} F_3(\mathbf{q}, -\mathbf{q}, \mathbf{k}) P_L(q), \end{aligned} \quad (2.40)$$

with F_3 being the kernel for the third order density field perturbation $\delta^{(3)}$ (see for example [54, 55] for the full expression). Diagrammatically, the above integrals can be represented

as follows

$$P_{22} \sim \begin{array}{c} P_L(q) \\ \circlearrowleft \\ \circlearrowright \\ P_L(|\mathbf{k}-\mathbf{q}|) \end{array} \quad P_{13} \sim \begin{array}{c} P_L(q) \\ \circlearrowleft \\ \circlearrowright \\ P_L(k) \\ \circlearrowleft \\ P_L(k-q) \end{array} \quad (2.41)$$

where the convention from [64] is used. Both integrals involve loops, but can nonetheless be safely numerically evaluated for the modes of interest, for example by a procedure shown in [65] (see their Chapter 2.2). However, the problems arise in the limiting cases for the momenta involved. For example, looking at the behavior of the $F_2(\mathbf{q}, \mathbf{k} - \mathbf{q})$ kernel in the case when $k \ll q$ one obtains $F_2 \sim \mathcal{O}(q/k)$, which suggests that there is an IR divergence. It turns out this effect is due to the presence of $v^i \partial_i \delta$ and $v^i \partial_i \theta$ source terms from Eq. (2.19) (see [64]) and corresponds to a uniform displacement of shorter-scale modes by a large-scale mode. This however is an unobservable effect due to the invariance of the original system of coupled PDEs from Eq. (2.19) to bulk flows. It can be therefore regulated by choosing a more suitable coordinate system, which in this case is the coordinate system along the fluid flow (see Section 4.2 in [64]). This can be demonstrated on the case of the $\delta^{(2)}$ term derived in Eq. (2.25). One can notice that the IR divergence is caused by the term $\sim \partial_i \delta_1 \frac{\partial_i}{\nabla^2} \delta_1$. Transforming this expression to Fourier space, we notice that it gives rise to a term scaling as $\sim \mathcal{O}(k_2/k_1)$. When $k_1 \ll k_2$ this term diverges. However, switching to the fluid frame, gets rid of this IR divergence. To see this, expand the density fluctuation up to 2nd order

$$\begin{aligned} \delta^{[2]}(\mathbf{x}, a) &= \delta^{(1)}(\mathbf{x}, a) + \delta^{(2)}(\mathbf{x}, a) \\ &= a\delta_1(\mathbf{x}) + a^2 \left(\frac{5}{7}\delta_1^2(\mathbf{x}) + \frac{2}{7} \left(\frac{\partial_i \partial_j}{\nabla^2} \delta_1(\mathbf{x}) \right)^2 + \partial_i \delta_1 \left(\frac{\partial_i}{\nabla^2} \delta_1(\mathbf{x}) \right) \right) \end{aligned} \quad (2.42)$$

and apply the coordinate shift to the fluid frame

$$\mathbf{x} \rightarrow \mathbf{x}_\#[\mathbf{x}, \eta; 0] = \mathbf{x} + \int_0^\eta d\eta' \mathbf{v}^{(1)}(\mathbf{x}, \eta'), \quad (2.43)$$

where $\mathbf{v}^{(1)}(\mathbf{x}, \eta') = -\mathcal{H} \frac{\partial_i}{\nabla^2} \delta^{(1)}(\mathbf{x}, \eta')$ is the first order solution for the velocity field in EdS. The notation $\mathbf{x}_\#[\mathbf{x}, \eta; 0]$ is just for keeping track of the particle along the fluid flow, meaning that the particle was at the beginning of the fluid flow line at conformal time $\eta = 0$, and ended up at the coordinate \mathbf{x} at time η which corresponds to the coordinate $\mathbf{x}_\#$ along the fluid flow. Applying now the shift $\Delta \mathbf{x}^{(1)} = \mathbf{x}_\# - \mathbf{x}$ to the terms on the right hand side of Eq. (2.42), and keeping all the relevant terms up to 2nd order in density fluctuations, produces (see for example [66, 67])

$$\delta^{[2]}(\mathbf{x}, a) = a\delta_1(\mathbf{x}_\#[\mathbf{x}, \eta; 0]) + a^2 \left(\frac{5}{7}\delta_1^2(\mathbf{x}_\#[\mathbf{x}, \eta; 0]) + \frac{2}{7} \left(\frac{\partial_i \partial_j}{\nabla^2} \delta_1^2(\mathbf{x}_\#[\mathbf{x}, \eta; 0]) \right) \right). \quad (2.44)$$

As one can see the IR divergent term is not present anymore, since it has cancelled with the term from the Taylor expansion of $\delta^{(1)}$ after applying the shift $\Delta\mathbf{x}^{(1)}$. This is actually a generic feature which also holds at higher orders [64, 66]. One more important thing to note is that the derivatives with respect to \mathbf{x} , employed in the Taylor expansion above, should be corrected for the discrepancy between \mathbf{x} and the fluid flow coordinate \mathbf{x}_fl . At leading order, the derivative with respect to \mathbf{x} and \mathbf{x}_fl coincide, while at higher orders one needs to account for the higher order terms coming from the Jacobian of this coordinate transform as well (see for example Appendix B). The intrinsic property of removing the IR divergent terms by simply modelling the evolution along the fluid flow, motivates our use of Lagrangian perturbation theory for the forward modelling effort from Chapter 3.

A direct consequence of this conclusion is that now we can claim that the time evolution diagram at n^{th} order is double soft. This can be also manifestly seen by introducing momentum density $\boldsymbol{\pi}^i = (1 + \delta)\mathbf{v}^i$, taking again the divergence of the Euler equation for the density perturbation from Eq. (2.17) and reducing it to second order ODE (see also Chapter 5.2 in [64]). This results in a source term, suppressing the arguments,

$$\begin{aligned} S_\delta &= \partial_i(\delta\partial_i\phi) + \partial_i\partial_j\left(\frac{\pi^i\pi^j}{1+\delta}\right) \\ &= \frac{2}{3\mathcal{H}^2}\left(\partial_i\partial_j(\partial_j\phi\partial_i\phi) - \frac{1}{2}\nabla^2(\partial_j\phi\partial_j\phi)\right) + \partial_i\partial_j\left(\frac{\pi^i\pi^j}{1+\delta}\right), \end{aligned} \quad (2.45)$$

where we have used the Poisson equation and product rule identities to go from the first to second line. As can be seen, all the source terms have two overall derivatives in front. This means that when solving for the density perturbation, all diagrams that appear will have the same overall scaling in derivatives, since the source term, S_δ , dictates what couplings are possible in the theory and Eq. (2.45) is non-perturbative. Therefore, since we know there are no IR divergences, i.e. no $\sim \frac{1}{\delta}$ terms are allowed in the interaction vertices, the overall double softness of the diagrams at all orders is preserved. If the IR safety was not confirmed beforehand, one could not claim double softness directly from Eq. (2.45).

We can now turn our attention towards looking at the UV limit of our theory. One can again analyze the behavior of the P_{22} and P_{13} from Eq. (2.40). In order to immediately see what is the scaling, one can write out the correlators as follows

$$P_{22} \sim \left\langle \begin{array}{c} \diagup \quad \diagdown \\ \diagdown \quad \diagup \end{array} \right\rangle, \quad P_{13} \sim \left\langle \begin{array}{c} | \\ \diagdown \quad \diagup \end{array} \right\rangle \quad (2.46)$$

and notice that the P_{22} arises from contracting two diagrams, each of which consist of two hard modes with momenta \mathbf{q} and $\mathbf{k} - \mathbf{q}$ (see also diagrams from Eq. (2.41)). The double softness argument implies that P_{22} has to scale as $\sim k^4$. The P_{13} contains the third order field $\delta^{(3)}$ which consists of one soft mode $\delta_1(\mathbf{k})$ and two hard modes, which is contracted

with the linear field $\delta^{(1)}$. This means it has to scale as $\sim k^2 P_L(k)$. The interpretation of the UV limit for the P_{22} comes directly from studying the constituent diagrams. As can be seen, both diagrams entering the correlator represent two hard modes combining into a soft mode. In other words, even in the absence of initial long mode perturbation, it will form from the collapse of the shorter scale modes. Given that a field with correlation structure scaling as $\sim k^4$ in Fourier space corresponds to a completely localized and uncorrelated field in real-space, it can be understood as stochastic noise contribution arising from short scales. On the other hand, the P_{13} arises from coupling an existing soft mode to hard modes. Therefore, the $\delta^{(3)}$ diagram appearing in the P_{13} correlator, describes how the evolution of short scale modes is affected in the presence of initial long mode. With this physical interpretation in mind, it is then possible to regulate the integrals from Eq. (2.40) by cutting them off at some momentum Λ and absorbing the residual cutoff dependent terms by adding appropriately scaling terms to the theory. These terms will capture the effective couplings between the remaining soft and neglected hard modes. In the case of 2-point function, given the above estimated scaling of the leading contributions at 1-loop, the leading order correction that needs to be included is

$$- 2c_{13}^2(\eta)k^2 P_L(k) + c_{22}^2(\eta)k^4, \quad (2.47)$$

with two new counter terms with time dependent coefficients appearing. They correspond to adding the following correction to the source term of the Euler equation

$$\delta_{\text{ct}}(\mathbf{k}, \eta) = c_{22}(\eta)k^2 \epsilon(\mathbf{k}) - c_{13}(\eta)k^2 \delta^{(1)}(\mathbf{k}, \eta), \quad (2.48)$$

with the introduction of $\epsilon(\mathbf{k})$ field, representing the stochastic noise contribution. This field is uncorrelated with the matter field δ , and has an analytic power spectrum, typically with increasing power on small scales (see for example [31]). The arbitrary time dependence of the counter term coefficient allows for systematic renormalization of all higher-order diagrams which include the same 1-loop contribution from the P_{22} and P_{13} diagrams. For example, analyzing the diagram for the 1-loop bispectrum contribution, B_{411}



$$B_{411} \sim \text{diagram}, \quad (2.49)$$

we see that the same 1-loop contribution as for P_{13} appears. However, the time dependence of the field ingoing into the loop is different, i.e. it is second order, arising from two linear fields coupling at the vertex labelled with “a”, while for the case of P_{13} it was first order. This is the reason why the most general coefficient of the counter term has to be time dependent. Note that the form of the leading order counter terms from Eq. (2.48) will stay the same even if we have assumed Λ CDM from the start, since the time- and spatial-dependent part also decouple as in EdS. Only the specific time dependence of c_{13}, c_{22} terms will be affected. This generic time dependence of the counter terms simply

reflects the fact that the short mode can be affected by a long mode at any moment during its evolution and then couple to other modes at some later time afterward (see for example Chapter 4.1 in [64]). Precisely the terms from Eq. (2.48) will be contained inside the neglected stress tensor τ^{ij} from Eq. (2.16), as will be shortly explained. Furthermore, this kind of analysis directly motivates the structure of our own EFTofLSS forward models from Chapter 3. For example, the $\epsilon(\mathbf{k})$ term directly dictates the k -dependence of the noise model from Section 3.2.3, while the second term, $c_{13}(\eta)k^2\delta^{(1)}(\mathbf{k}, \eta)$, motivated the presence of $b_{\nabla^2\delta}$ term in Eq. (3.11). Although, see Section 2.4 for a more complete description.

In order to systematically generate required counter terms at all orders, it is enough to write out, at the given order, all terms consistent with the equivalence principle, which are as well invariant to time-dependent boosts. This suggests that the terms will start first order in δ but second order in derivatives. All the terms can then be absorbed into the effective stress tensor, τ^{ij} , from Eq. (2.16). For example, the $\sim \delta^{(1)}$ term from Eq. (2.48) is obtained from the following effective stress tensor

$$\tau^{ij} \supset l_1(\eta)\partial^i\partial^j\Psi + l_2(\eta)\delta_D^{ij}\partial^2\Psi, \quad (2.50)$$

while at next-to-leading order, relevant for the bispectrum for example, we have

$$\tau^{ij} \supset p_1(\eta)\partial^i\partial_k\Psi\partial^k\partial^j\Psi + p_2(\eta)\partial^2\Psi\partial^2\Psi + p_3(\eta)\partial^i\partial^j\Psi\partial^2\Psi. \quad (2.51)$$

The l_1, l_2, p_1, p_2 and p_3 are all generic time dependent functions. The above discussion focused on density perturbations, δ , however the same approach holds for the θ field as well, given that it can be associated with velocity potential $\Psi_v \equiv \nabla^{-2}(\frac{1}{H}\theta)$. For consistent treatment of both δ and θ field solutions at all orders, see [64].

The equations 2.50 and 2.51 can be written in a way where the invariance to the time dependent boosts is more manifest as well as the preservation of IR safety of the theory. In this form the expression for τ^{ij} is given as (see Eq. 6.9 in [64])

$$\tau^{ij}(\mathbf{x}, \eta) = \int_0^\eta \{d\eta_n\} K_{a_1\dots a_n\dots}(\eta; \{\eta_m\}) \underbrace{\Pi_{a_1}^{i_1j_1} \dots \Pi_{a_k}^{i_kj_k} \dots \Pi_{a_l}^{i_lj_l} \dots}_n, \quad (2.52)$$

where

$$\Pi_a^{ij} \equiv \Pi_a^{ij}(\mathbf{x}_\#(\eta_a), \eta'_a) \stackrel{\text{def}}{=} \partial^i\partial^j\Psi_a(\mathbf{x}(\eta_a), \eta'_a),$$

with Ψ_a representing a doublet of density and velocity potentials

$$\Psi_0 = \Psi, \quad \Psi_1 = \Psi_v \quad (2.53)$$

As can be seen they are evaluated along the fluid flow and are manifestly IR safe as well as invariant to time dependent boosts. Furthermore, these terms are all observables, since they contain two spatial derivatives, and hence consistent with the equivalence

principle. The $K_{a_1 \dots a_n}(\eta; \{\eta_n\})$ represents an arbitrary time dependent kernel, reflecting the arbitrary time dependence of an ingoing field to the loop at n^{th} order. As an example, one can consider the terms from Eq. (2.50), apply two derivatives to the Euler equation from Eq. (2.16), in order to obtain the source term for the velocity divergence PDE. This then directly results in a term

$$\begin{aligned} \frac{1}{\rho} \partial_i \partial_j \tau^{ij} &\sim \partial_i \partial_j \tau^{ij} + \mathcal{O}(\delta^2) \\ &\supset \partial_i \partial_j \int_0^\eta d\eta' K_{11}(\eta; \eta') [l_1(\eta') \partial^i \partial^j \Psi_1(\mathbf{x}_\text{fl}, \eta') + l_2(\eta') \delta_D^{ij} \partial^2 \Psi_1(\mathbf{x}_\text{fl}, \eta')] \\ &= \partial^2 \int_0^\eta d\eta' K_{11}(\eta; \eta') (l_1(\eta') + l_2(\eta')) \delta(\mathbf{x}_\text{fl}, \eta'), \end{aligned} \quad (2.54)$$

which already reproduces the form of the $\delta^{(1)}$ from Eq. (2.50). The above expression is satisfying both the equivalence principle requirements and invariance under time dependent boosts, given that it is formulated along the fluid flow. In order to see what is the relation between the c_{13} from Eq. (2.50) and the coefficients l_1, l_2 one can just perturbatively solve for $\delta(\mathbf{x}_\text{fl}(\eta), \eta')$ in terms of $\delta(\mathbf{x}, \eta)$ (see for example Chapter 3.1 in [67]) to obtain

$$\delta(\mathbf{x}_\text{fl}(\eta), \eta') = \frac{a(\eta')}{a(\eta)} \delta(\mathbf{x}, \eta) + \frac{a(\eta')}{a(\eta)} \left(\frac{a(\eta')}{a(\eta)} - 1 \right) \left[\frac{5}{7} \delta^2(\mathbf{x}, \eta) + \frac{2}{7} \left(\frac{\partial^i \partial^j}{\nabla^2} \delta(\mathbf{x}, \eta) \right)^2 \right] + \mathcal{O}(\delta^3) \quad (2.55)$$

which holds up to second order in δ . One can then easily see that, after applying the usual perturbative ansatz from Eq. (2.28), the relation is

$$c_{13}(\eta) = \int_0^\eta d\eta' K'_{11}(\eta, \eta') (l_1(\eta') + l_2(\eta')), \quad (2.56)$$

where $K \rightarrow K'$ emphasizes that the time-dependent part of the perturbative solution can just be absorbed by the kernel itself. Note that this result will also hold in general Λ CDM case too. Therefore, the generic relation from Eq. (2.52) is indeed capable of reproducing the $\sim \delta^{(1)}$ term from Eq. (2.48) as anticipated, and can be used for generating higher order counter terms. The only remaining piece to be explained is the generation of stochastic terms, such as the $\sim \epsilon(\mathbf{k})$ in Eq. (2.48). However, these terms are subleading on large scales. This can already be suspected from the scaling in Eq. (2.47), given that the amplitude of the stochastic and non-stochastic contribution scale as

$$\Delta_{c_{22}}^2 \sim k^7, \quad \Delta_{c_{13}}^2 \sim k^{(5+n_s)}, \quad (2.57)$$

where n_s is the slope of the linear power-spectrum close to k_{NL} , which is approximately $n_s \sim -1.5$, since the typical value for $k_{\text{NL}} \sim 0.3 h \text{ Mpc}^{-1}$ (see discussion below Eq. (2.38)).

Given that we will be concerned only with the leading order piece of the stochasticity, as given in Eq. (2.48) in Chapter 3, we refer the reader to consult with [31] (Chapter 2.8) for a systematic treatment of the higher order stochastic terms.

The momentum cutoff of the loops from Eq. (2.41), necessary for regulating the SPT approach, can be also generically imposed if one would smooth the original equations of motion, Eq. (2.16), up to the same scale. This will automatically regulate the loops up to any order. Even if one would start from the approximate equations from 2.17, smoothing will introduce an effective stress-tensor as a source term. Therefore, the appearance of the effective stress-tensor is the generic feature of the matter clustering problem. It simply comes from the fact that the gravitational interaction couples all scales. The smoothed equations Eq. (2.16) take the form

$$\begin{aligned}\delta'_l + \partial_i \left((1 + \delta_l) v_l^i \right) &= 0, \\ v_l^{i'} + v_l^j \partial_j v_l^i + \mathcal{H} v_l^i + \partial^i \Psi_l &= -\frac{1}{\rho_l} \partial_j [\tau^{ij}]_\Lambda, \\ \nabla^2 \Psi_l &= \frac{3}{2} \mathcal{H}^2 \Omega_m(\eta) \delta_l,\end{aligned}\tag{2.58}$$

where subscript $(\cdot)_l$ denotes long modes that remain after convolving with the smoothing kernel with a characteristic scale Λ . In order to derive the above result, one should start directly smoothing the momenta involved in the original Boltzmann equation Eq. (2.15) (see [27] for derivation). The expression of the effective stress tensor arising from this smoothing procedure is given as

$$[\tau^{ij}]_\Lambda = [\rho v^i v^j]_\Lambda + \frac{1}{8\pi G} [2\partial_i \Psi_s \partial_j \Psi_s - \delta_D^{ij} (\nabla \Psi_s)^2]_\Lambda,\tag{2.59}$$

where the superscript $(\cdot)_s$ stands for short modes, i.e. the ones with $k \geq \Lambda$. Given that the curl-free velocity field can be written as the gradient of a velocity potential field, all the terms appearing above have two spatial derivatives, as required by the equivalence principle.

Another important thing to note is the justification for the truncation of the Boltzmann hierarchy at the second moment. To really label the dark matter content of the universe as a cosmic fluid requires it not only to respect continuity and Euler equations from Eq. (2.16), but also have negligible higher order moments of the Boltzmann equation. For neutrinos this was demonstrated in [52], while in similar manner, reference [27] shows that the same conclusion holds for cold dark matter. Namely, after writing the recursion relation of the Boltzmann equations for the perturbation of the dark matter phase space distribution function for moments $m \geq 2$ (using the same convention as in [52])

$$\dot{\Delta}_f^{[m]}(\mathbf{k}, \eta) = \frac{k|\dot{\mathbf{x}}|}{2m+1} \left(m \Delta_f^{[m-1]}(\mathbf{k}, \eta) - (m+1) \Delta_f^{[m+1]}(\mathbf{k}, \eta) \right), \quad m \geq 2,\tag{2.60}$$

where $v_p \stackrel{\text{def}}{=} |\dot{\mathbf{x}}|$ is the particle's peculiar velocity, which doesn't necessarily coincide with the mean peculiar velocity obtained from calculating moments of the Boltzmann equation (see Eq. (2.16)). The relevant timescale is the Hubble time, $\sim \mathcal{H}^{-1}$, and hence the solution for the Δ_f can be roughly estimated to scale as

$$\Delta_f^{[m]}(\mathbf{k}, \eta) \sim \frac{kv_p \mathcal{H}^{-1}}{2m+1} \left(m \Delta_f^{[m-1]}(\mathbf{k}, \eta) - (m+1) \Delta_f^{[m+1]}(\mathbf{k}, \eta) \right), \quad m \geq 2. \quad (2.61)$$

From the above it follows that

$$\Delta_f^{[m]} \sim (kv_p \mathcal{H}^{-1})^{m-2} \Delta_f^{[2]}. \quad (2.62)$$

This is similar as to the usual fluid treatment, where the prefactor is instead $\sim kv_p t_c$, with t_c being the characteristic collision time. As usually done, one coarse grains the Boltzmann hierarchy, takes volume elements of big enough extent such that $kv_p t_c \ll 1$. In other words, one models only the phenomenology at scales $k \ll (v_p t_c)^{-1}$, much bigger than the characteristic length scale covered within the collision time. Unlike for the conventional fluid, where the characteristic collision time is very short, the dark matter fluid has a characteristic timescale of order $\sim \mathcal{H}^{-1}$. Since we're interested in the modes within the horizon, it follows that the characteristic magnitude of the dark matter peculiar velocities should be small in order to allow for the truncation of Boltzmann hierarchy. To see this, one can look at the typical amplitude of the velocity field and see when it becomes $\mathcal{O}(1)$

$$\Delta_v^2(k_{\text{NL}}) \sim \Delta_\delta^2(k_{\text{NL}}) \left(\frac{\mathcal{H}}{k_{\text{NL}}} \right)^2 \sim \left(\frac{\mathcal{H}}{k_{\text{NL}}} \right)^2, \quad (2.63)$$

since $\Delta_\delta(k_{\text{NL}}) \sim 1$, and the relation $v_p(\mathbf{k}, \eta) \sim k^{-1} \mathcal{H} \delta(\mathbf{k}, \eta)$ follows directly from the leading order continuity equation (see Eq. (2.17)). Therefore, we see that as long as

$$kv_p \mathcal{H}^{-1} \lesssim \frac{k}{k_{\text{NL}}}, \quad (2.64)$$

i.e. $k \ll k_{\text{NL}}$, the truncation of the Boltzmann hierarchy is well motivated and hence dark matter can be indeed treated as an effective fluid.

At the end of the day, we will not be able to directly observe the dark matter field, or the halos that form by the gravitational evolution, as described in this section. Instead, we observe a tracer of the matter field. This tracer can be anything from galaxies, quasars, the Ly- α forest [68, 69], Sunyaev-Zel'dovich effect measurements [3, 70–72], 21cm observations [73], line intensity mapping data [74] etc. In this work we are focusing on galaxies, however the same principles will apply to any other tracer. In Section 2.4 we aim to provide a general framework of how the EFT approach can be extended to the description of biased tracers.

2.4 Modelling biased tracers

The previous sections were focused on the description of modelling the underlying matter density field. However, we can not directly observe dark matter, rather it is much easier to deal with different tracers of this underlying matter field observationally, such as galaxies, quasars etc. All these tracers are collectively known as *biased tracers* of the large-scale structure. The main goal of all large-scale structure models is to constrain cosmological parameters given the biased tracer data. Therefore, we dedicate this section to the effective theory of biased tracers, upon which the section Chapter 3 relies.

The most general bias expansion follows from the following assumptions: (i) the Equivalence principle, stating that all objects feel the gravitational field in the same way and (ii) the renormalization condition, saying that one needs to include in the bias expansion all terms allowed by the symmetries of general relativity up to the given order. The first assumption suggests that there must be a mapping between the underlying matter field and the clustered tracer field (halos or galaxies for example), i.e.

$$\delta_g(\mathbf{x}, \tau) = \sum_{\mathcal{O}} b_{\mathcal{O}}(\tau) \mathcal{O}[\delta, \Theta](\mathbf{x}, \tau), \quad (2.65)$$

where we have denoted with the $\mathcal{O}[\delta, \Theta](\mathbf{x}, \tau)$ a general bias operator, which itself is a function of the underlying matter field δ and cosmology Θ , while $b_{\mathcal{O}}(\tau)$ are the corresponding bias coefficients. The time dependence of the bias coefficients reflects the fact that different sized objects will collapse at different times, as already noted in Section 2.3. The leading order term can already be fixed through the famous *peak-background split* argument, first proposed by N. Kaiser in 1984 [75]

$$\delta_g(\mathbf{x}, \tau) \supset b_{\delta}(\tau) \delta(\mathbf{x}, \tau) \quad (2.66)$$

Although this original argument was derived for dark matter halos, the same argument applies between the halo and galaxy density field due to the equivalence principle. Hence, by transitivity, the same relation will hold between the matter and galaxy clustering field as written above. Since the voids are as much present in our universe as clusters of galaxies, the bias coefficient b_{δ} can be any real number. Following this logic, one can then propose the following bias expansion in a form of a Taylor series [76, 77]

$$\delta_g^{\text{LIMD}}(\mathbf{x}, \tau) = \sum_n b_n(\tau) \delta^n[\Theta](\mathbf{x}, \tau). \quad (2.67)$$

Given that we are forming a perturbation series description, it is necessary to consider overdensities which are below $\mathcal{O}(1)$. The same reasoning as from Section 2.3 then suggests that we need to impose a scale up to which we want to describe the galaxy overdensity field, given that the smaller the scales we are concerned with, the larger the amplitude of

the overdensity and hence weaken the PT description. However, a Lagrangian approach will extend the applicability of the above expansion to beyond $\mathcal{O}(1)$ overdensities (see Appendix B and Chapter 3). In the spirit of the perturbative description from Section 2.3 and 2.1, the above equation should be rephrased as

$$\text{Eq. (2.67)} \rightarrow \delta_{g,\Lambda}^{\text{LIMD}}(\mathbf{x}, \tau) = \sum_n b_n(\tau) \delta_\Lambda^n[\Theta](\mathbf{x}, \tau), \quad (2.68)$$

with Λ denoting the cutoff scale. This approach is known as *local-in-matter-density* (LIMD) bias expansion (see Section 2.1 in [31]), hence the superscript. As the name suggests, the expansion doesn't account for spatially non-local effects, such as tidal field for example, which is a local observable. This already suggests the above expansion is incomplete, since it is expected the tracer density should also depend on the underlying tidal field. In order to see what is necessary how to expand the above PT series, one can start from looking at the expectation value of the proposed LIMD mapping. Given that we are interested in modelling an overdensity field we need to have

$$\langle \delta_{g,\Lambda}^{\text{LIMD}} \rangle \stackrel{!}{=} 0, \quad (2.69)$$

however, already with $n = 2$ in Eq. (2.67) we have (suppressing arguments)

$$\begin{aligned} \langle \delta_{g,\Lambda}^{[2],\text{LIMD}} \rangle &= \langle b_\delta \delta_\Lambda^{[2]} + b_{\delta^2} (\delta_\Lambda^{(1)})^2 \rangle = \langle (\delta_\Lambda^{(1)})^2 \rangle \\ &= \int_{\mathbf{q}}^\Lambda P_L(q) \equiv \sigma^2(\Lambda), \end{aligned} \quad (2.70)$$

given that $\langle \delta_\Lambda^{[2]} \rangle = 0$. We see therefore that the mean of the PT expansion up to 2nd order is not zero, but equals the variance of the matter field, smoothed on scale Λ . The issue can be alleviated by renormalizing the expansion and introducing a counter term at 2nd order as

$$\delta_\Lambda^{[2]} = b_\delta \delta_\Lambda + b_{\delta^2} [\delta_\Lambda^2] = b_\delta \delta_\Lambda + b_{\delta^2} ((\delta_\Lambda)^2 - \sigma^2(\Lambda)). \quad (2.71)$$

As can be explicitly checked, this renormalized 2nd order bias expansion now correctly produces zero mean expectation value for the resulting galaxy overdensity field. Since we are not only interested in the 1-point function, but want to also model the higher order moments of the galaxy field, we need to also ensure the final result doesn't depend on the arbitrary selected cutoff Λ . The same issue was already encountered when considering the 2-point function of the matter field in Section 2.3 which resulted in the introduction of counter terms of the form Eq. (2.48) at the one loop level. Let us therefore consider the 2-point function of the galaxy clustering field, at one loop. In order to grasp the main argument, there is no need to go beyond δ^2 in bias expansion. With this, we have

$$\begin{aligned} \left\langle \delta_{g,\Lambda}^{[2],\text{LIMD}} \delta_{g,\Lambda}^{[2],\text{LIMD}} \right\rangle'_{1\text{-loop}} &= b_\delta^2 \langle \delta_\Lambda \delta_\Lambda \rangle'_{1\text{-loop}} \\ &+ 2b_\delta b_{\delta^2} \langle (\delta_\Lambda)^2 \delta_\Lambda \rangle'_{1\text{-loop}} + b_{\delta^2}^2 \langle (\delta_\Lambda)^2 (\delta_\Lambda)^2 \rangle'_{1\text{-loop}}, \end{aligned} \quad (2.72)$$

as usual, ' denotes dropping the momentum conserving Dirac deltas. The first term from Eq. (2.72) represents the two point function of the matter field, and is renormalized within the EFTofLSS as demonstrated in Section 2.3 at 1-loop. However, as we will see shortly, the terms in the second line require a new type of counter terms, different from those introduced for the matter field in Eq. (2.48) at the 1-loop level. In order to see this, we can expand the $\sim b_\delta b_{\delta^2}$ and $\sim b_{\delta^2}^2$ terms in density perturbations up to $\sim (\delta^{(1)})^4$

$$\begin{aligned}
\text{Eq. (2.72)} \supset & \left\langle \left(\delta_\Lambda^{(1)} + \delta_\Lambda^{(2)} \right)_q^2 \left(\delta_\Lambda^{(1)} + \delta_\Lambda^{(2)} \right)_{q_1} \right\rangle' \\
& + \left\langle \left(\delta_\Lambda^{(1)} \right)_q^2 \left(\delta_\Lambda^{(1)} \right)_{q_1}^2 \right\rangle' \\
& = \left\langle \left(\delta_\Lambda^{(1)} \delta_\Lambda^{(1)} \right)_q \left(\delta_\Lambda^{(2)} \right)_{q_1} \right\rangle' + 2 \left\langle \left(\delta_\Lambda^{(1)} \delta_\Lambda^{(2)} \right)_q \left(\delta_\Lambda^{(1)} \right)_{q_1} \right\rangle' \\
& + \left\langle \left(\delta_\Lambda^{(1)} \right)_q^2 \left(\delta_\Lambda^{(1)} \right)_{q_1}^2 \right\rangle', \tag{2.73}
\end{aligned}$$

or diagrammatically

$$\begin{aligned}
\text{Eq. (2.72)} \supset & \begin{array}{c} \text{Diagram 1} + \text{Diagram 2} \\ + \text{Diagram 3} \end{array}, \tag{2.74}
\end{aligned}$$

with the white circle representing convolution in Fourier space. In the UV limit, each loop term evaluates to (schematically)

$$\begin{aligned}
\text{Diagram 1} & \sim \int_{\mathbf{p}}^{\Lambda} F_2(\mathbf{p}, \mathbf{q} - \mathbf{p}) P_L(|\mathbf{q} - \mathbf{p}|) P_L(p) \\
& \sim q^2 \left(\int_{\mathbf{p}}^{\Lambda} \frac{dp}{2\pi^2} P_L^2(p) \right) \\
\text{Diagram 2} & \sim \sigma^2(\Lambda) P_L(q) \\
\text{Diagram 3} & \sim \sigma^4(\Lambda). \tag{2.75}
\end{aligned}$$

The first line scales as q^2 , which follows directly from the behavior of the F_2 kernel for $q \ll p$, i.e. $\lim_{q \ll p} F_2(\mathbf{p}, \mathbf{q} - \mathbf{p}) \rightarrow q^2/p^2$. Borrowing from the intuition gathered

while analyzing the P_{22} loop term from Eq. (2.41), we can conclude that this term will contribute as a stochastic noise term with correlation structure scaling as $\sim q^2$ in Fourier space. This demonstrates that the stochasticity of biased tracers has a different structure to the one of the matter field and requests a term with correlation structure $\sim q^2$ to be as well inserted.

Similarly, the other two diagrams we have also not encountered in previous section at the level of 1-loop matter power spectrum. As we can see, these loops arise from contracting the fields within the $\sim \delta^2$ bias operator. This insight will help us formulate a more general renormalization condition, which will provide a systematic way of generating all needed counter terms at the given loop order (see Eq. (2.85)). This will in turn generalize the bias expansion from Eq. (2.67) and force us to include all possible terms allowed by the symmetry of the problem at the given loop order, a recurring theme of every EFT approach.

It is therefore instructive to analyze more closely the remaining two terms. Looking back at the Eq. (2.71) it is clear that the $\sim \sigma^4(\Lambda)$ term of Eq. (2.75), will be automatically cancelled due to the introduced $\sim \sigma^2(\Lambda)$ term in the renormalized expression for δ^2 . However, the $\sim \sigma^2(\Lambda)P_L$ requires an introduction of yet another counter term, this time of the form $\sim \sigma^2(\Lambda)\delta_\Lambda$. In order to see exactly what is the term to add, we can explicitly evaluate the loop

$$\begin{aligned} \langle (\delta_\Lambda)^2(\mathbf{q})\delta_\Lambda(\mathbf{q}_1) \rangle'_{1\text{-loop}} &= 2 \int_{\mathbf{p}_1} \langle \delta_\Lambda^{(1)}(\mathbf{p}_1)\delta_\Lambda^{(2)}(\mathbf{q}-\mathbf{p}_1)\delta_\Lambda(\mathbf{q}_1) \rangle' \\ &= 2W_\Lambda(\mathbf{q}_1) \int_{\mathbf{p}_1} \int_{\mathbf{p}_2} \int_{\mathbf{p}_3} (2\pi)^3 \delta_D(\mathbf{q}-\mathbf{p}_1-\mathbf{p}_2-\mathbf{p}_3) \\ &\quad W_\Lambda(p_1)W_\Lambda(|\mathbf{p}_2+\mathbf{p}_3|)F_2(\mathbf{p}_2,\mathbf{p}_3) \\ &\quad \langle \delta^{(1)}(\mathbf{p}_1)\delta^{(1)}(\mathbf{p}_2)\delta^{(1)}(\mathbf{p}_3)\delta^{(1)}(\mathbf{q}_1) \rangle' \\ &= 4W_\Lambda(\mathbf{q}_1) \left[\int_{\mathbf{p}} W_\Lambda(p)W_\Lambda(|\mathbf{q}+\mathbf{p}|)F_2(\mathbf{q},\mathbf{p})P_L(p) \right] P_L(q_1), \end{aligned}$$

dropping momentum conserving Dirac deltas. The first line expands the convolution from the $(\delta_\Lambda)^2$ operator, while the next one applies the solution for the $\delta^{(2)}$ from Eq. (2.26). The kernels W_Λ are representing the smoothing operation, and cut the corresponding momenta up to Λ in magnitude. Divergence generated in the limit of $q \ll q_1$ is handled by our discussion on EFTofLSS from previous section (Section 2.3), the more interesting limit here is when $q_1 \ll q$. Therefore, we can take $W_\Lambda(\mathbf{q}_1) \rightarrow 1$. Similarly, taking $q \ll \Lambda$, since we're interested in the UV behavior of the loop, one can perform the angular integration of the momentum integral and expand in powers of (q_1/q) to obtain

$$\langle (\delta_\Lambda)^2(\mathbf{q})\delta_\Lambda(\mathbf{q}_1) \rangle'_{1\text{-loop}} = 4P_L(q_1) \left(\frac{17}{21} \int_{\mathbf{p}} W_\Lambda(p)P_L(p) \right) = \frac{68}{21}\sigma^2(\Lambda)P_L(q_1). \quad (2.76)$$

This then suggests the bias operator δ^2 should be added a counter term of the form

$$[\delta^2] = \delta_\Lambda^2 - \sigma^2(\Lambda) \left(1 + \frac{68}{21} \delta_\Lambda \right). \quad (2.77)$$

The correction doesn't yet suggest that the original expansion from Eq. (2.67) is incomplete, given that one can redefine the leading order bias coefficient $b_\delta \rightarrow b_\delta - (68/21) \sigma^2(\Lambda)$ and absorb the divergence. This indicates that the bias parameters themselves as given by the LIMD expression from Eq. (2.67), in this case b_δ , are going to depend on the cut-off scale, and therefore can not be seen as observable quantities. It is however possible to renormalize the bias parameters as well. As shown in [78], it follows automatically after listing all possible terms allowed by the equivalence principle at the given order in density fluctuations and thus providing a complete bias expansion at that order.

To see how to systematically generate a complete bias expansion, it is instructive to proceed to the 1-loop expansion from Eq. (2.72) even further, up to $\sim (\delta^{(1)})^6$. In order not to worry about the terms renormalized by the EFTofLSS approach from Section 2.3, it is better to focus only on the 1PI⁸ diagrams appearing at 1-loop, these are diagrammatically (see also [78])

$$\left\langle \delta_{g,\Lambda}^{(n=2)} \delta_{g,\Lambda}^{(n=2)} \right\rangle_{1\text{-loop}}^{\text{1PI}} = \text{Diagram 1} + \text{Diagram 2} + \text{Diagram 3}$$

The first diagram we have already seen in Eq. (2.75), with the only difference here being that we should keep in mind it appears as a part of a larger diagram whose remaining part is renormalized by the EFTofLSS counter terms. The structure of this 1PI diagram immediately suggests that the corresponding counter term should scale as $\sim \sigma^2 \delta_\Lambda$ as we already discussed. The diagrams on the second line are new.

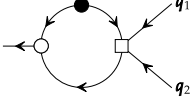
Following the same logic, we can then immediately write down what is the scaling of the first 1PI diagram on the second line with Λ

$$\text{Diagram 2} \sim \left(\int^{\Lambda} \mathbf{p} F_2(\mathbf{p}, \mathbf{q}_1) F_2(\mathbf{p}, -\mathbf{q}_2) P_L(p) \right) \delta(\mathbf{q}_1) \delta(\mathbf{q}_2), \quad (2.78)$$

which can be absorbed by adding a counter term proportional to $\sim \delta^2$ to the renormalization of $\delta^{[2]}$ bias operator. The Λ dependence of the remaining diagram however can

⁸one-particle irreducible

not. The resulting expression of this diagram is



$$\sim \left(\int_{\mathbf{p}}^{\Lambda} F_3(\mathbf{q}_1 + \mathbf{q}_2 - \mathbf{p}, \mathbf{q}_1, \mathbf{q}_2) P_L(p) \right) \delta_{\Lambda}(\mathbf{q}_1) \delta_{\Lambda}(\mathbf{q}_2), \quad (2.79)$$

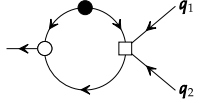
where the white square vertex symbolizes the F_3 kernel given by (see e.g. [54, 55])

$$F_3(\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3) = \frac{1}{18} [7\alpha(\mathbf{q}_1, \mathbf{q}_2 + \mathbf{q}_3) F_2(\mathbf{q}_2, \mathbf{q}_3) + 2\beta(\mathbf{q}_1, \mathbf{q}_2 + \mathbf{q}_3) G_2(\mathbf{q}_2, \mathbf{q}_3)] \\ + \frac{1}{18} [7\alpha(\mathbf{q}_1 + \mathbf{q}_2, \mathbf{q}_3) + 2\beta(\mathbf{q}_1 + \mathbf{q}_2, \mathbf{q}_3)] G_2(\mathbf{q}_1, \mathbf{q}_2),$$

with

$$\alpha(\mathbf{q}_1, \mathbf{q}_2) = 2F_2(\mathbf{q}_1, \mathbf{q}_2) - G_2(\mathbf{q}_1, \mathbf{q}_2), \\ \beta(\mathbf{q}_1, \mathbf{q}_2) = \frac{5}{2}G_2(\mathbf{q}_1, \mathbf{q}_2) - \frac{3}{2}F_2(\mathbf{q}_1, \mathbf{q}_2). \quad (2.80)$$

Taking the limit of large loop momentum $q_1, q_2 \ll p$ and performing the angular integral after expanding the F_3 kernel gives (see also [78])



$$\sim \frac{(\mathbf{q}_1 \cdot \mathbf{q}_2)^2}{q_1^2 q_2^2} \delta_{\Lambda}(\mathbf{q}_1) \delta_{\Lambda}(\mathbf{q}_2) + \mathcal{O}\left(\frac{q_i^2}{\Lambda^2}\right), \quad (2.81)$$

which suggests a completely new counter term, constructed out of the 2nd derivative of the gravitational potential $\partial^2 \Psi_{\Lambda}$ is to be included in order to properly renormalize δ^2 at the level of 1-loop. As expected, this term is indeed a local observable and as such suggests that the original expansion in Eq. (2.67) is incomplete. More precisely, the relevant counter terms we looked at up to now renormalize the δ^2 bias operator as (see [78] and Section 2.10 in [31])

$$[\delta^2] = \delta_{\Lambda}^2 - \sigma^2(\Lambda) \left(1 + \frac{68}{21} \delta_{\Lambda} + \frac{2624}{735} \delta_{\Lambda}^2 + \frac{254}{2205} \left(\frac{2}{3\mathcal{H}^2 \Omega_m} \partial_i \partial_j \Psi_{\Lambda} \right)^2 \right). \quad (2.82)$$

In fact, the term $(\partial^2 \Psi_{\Lambda})^2$, can be rewritten in terms of the more familiar tidal tensor

$$K_{ij,\Lambda} = \frac{2}{3\mathcal{H}^2 \Omega_m} \partial_i \partial_j \Psi_{\Lambda} - \frac{1}{3} \delta_{ij} \delta_{\Lambda}, \quad (2.83)$$

which then adapts Eq. (2.82) to

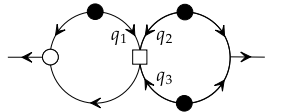
$$[\delta^2] = \delta_{\Lambda}^2 - \sigma^2(\Lambda) \left(1 + \frac{68}{21} \delta_{\Lambda} + \frac{24032}{6615} \delta_{\Lambda}^2 + \frac{254}{2205} (K_{ij,\Lambda})^2 \right). \quad (2.84)$$

Appearance of the tidal tensor in the bias expansion at 2nd order motivates the question on whether it itself is already renormalized or one needs additional terms in order to achieve that. As it turns out, by repeating the same reasoning as we have done so far, the necessary term for renormalizing the tidal tensor, whilst keeping up to the 2nd order in density perturbations is given as $\sim \partial^2 \delta$. This could've been also guessed directly from the form of the renormalized δ^2 operator from Eq. (2.84), since δ^2 appears directly in K_Λ^2 with two additional derivatives in front. This is precisely the reason why we include bias terms such as $\sim b_{\nabla^2 \delta}$ in our forward models from Chapter 3.

The intuition gathered so far can be formulated into a general *renormalization condition* for the bias operators as ([31, 78])

$$\lim_{k_i \rightarrow 0} \langle [O](\mathbf{k}) \delta^{(1)}(\mathbf{k}_1) \dots \delta^{(1)}(\mathbf{k}_n) \rangle \stackrel{!}{=} \langle O(\mathbf{k}) \delta^{(1)}(\mathbf{k}_1) \dots \delta^{(1)}(\mathbf{k}_n) \rangle_{\text{tree}}. \quad (2.85)$$

In words, we require that in the large-scale limit, all n -point functions the bias operator is involved in, don't have any diverging loop contributions. The reason why this condition is phrased through only contracting with n instances of the linear matter field $\delta^{(1)}$ is that the divergences specific to the biased tracer expansion are coming from the UV behavior of the bias operator O itself, and not from internal contractions between the constituent fields of the bias operator and higher order matter fields. These are renormalized by the approach presented in Section 2.3. For example, at the level of 2-loops the tracer power spectrum contains

$$\langle \delta_{g,\Lambda} \delta_{g,\Lambda} \rangle_{2\text{-loop}} \supset \text{diagram}, \quad (2.86)$$


However, the behavior in the limit $q_1 \ll q_2, q_3$ is already captured by the EFTofLSS counter terms. On the other hand, when $q_2, q_3 \ll q_1$, the UV behavior will match the one from the 1PI diagram in Eq. (2.79). Therefore, even though this term should formally be renormalized with the counter term appearing at two loops, the fact that we have already renormalized properly the matter density field allows us to focus only on the 1PI diagram coming from internal contractions of the corresponding bias operator.

In conclusion, if one is interested in producing a complete bias expansion at order n , one needs to renormalize all bias operators appearing to that order by applying the above renormalization condition for all possible number of matter fields $< n$. This ensures all loop divergences are captured by the generated counter terms. In fact, by this procedure, the bias parameters themselves get renormalized as well and don't depend on the smoothing scale Λ (see Section 3.2 in [78]). This then allows for physical interpretation of the obtained values of the bias parameters from any inference approach. As will be seen, in Chapter 3 we precisely use this fact to both construct, and later discuss, the inferred values of various bias parameters of the forward models we consider.

With this, one can then continue to any order in perturbation theory, which then produces a general bias expansion of the form

$$\delta_g(\mathbf{x}, \tau) = \sum_O b_O(\tau)[O](\mathbf{x}, \tau), \quad (2.87)$$

which is exactly what we used in Chapter 3 for our EFTofLSS forward model.

The first look into the performance of the EFTofLSS approach at the level of 2-point, for the matter density field at redshift zero was performed in [29], while for the halo density field it was first done in [79, 80]. As they demonstrated, the EFTofLSS approach is capable of reproducing the non-linear power spectrum at the percent level up to $k_{\text{NL}} \sim 0.3 h \text{Mpc}^{-1}$ and therefore significantly improves the results of the SPT approach as previously demonstrated for example in [62, 63]. The EFTofLSS approach however introduces more parameters, as any EFT does, which need to be calibrated for the application at hand. Examples of such parameters are bias parameters presented in this section or EFTofLSS counter terms from Section 2.3. What the authors of [29, 79, 80] did is to deduce them directly from the N-body simulation and then use them in their EFT expansion. Of course, since we don't have access to the N-body simulation for our own universe, one needs to marginalize over these parameters statistically by using some of the techniques explained in Section 2.5. This is precisely what was done in the works [14, 15], which were able to utilize the EFTofLSS approach and apply it at the level of 2- and 3-point functions on the 12th data release of the BOSS galaxy survey [81] and constrain cosmological parameters of the Λ CDM model using galaxy data for the first time. The question remains, however, how much information can be gained on the cosmology by going to ever higher n -point functions. One way to answer this question is to construct a field-level forward model which will automatically contain the information of all higher order n -point statistics. As briefly mentioned previously, and as elaborated in Section 2.6, this has its own downfalls. However, it remains to be demonstrated whether combining the field-level forward modelling approach with the EFTofLSS, taking the best of both worlds, is the most optimal way forward. A step towards this goal was exactly one of the aims of this thesis and is undertaken in Chapter 3. We will come back to this topic in Section 2.6.2.

2.5 Statistical inference

Before continuing onto the formulation of the field-level forward models of the large-scale structure, it is instructive to briefly review the statistical inference machinery employed throughout this thesis. This will allow for better understanding of the different model choices presented in Section 2.6. A common theme among the models is that all of them are built on top of Bayesian statistics, which we motivate below.

Bayesian reasoning arises as a natural formulation for problems where one needs to draw conclusions from incomplete information at hand. In 1946, R.T Cox [82] showed that there is unique correspondence between reasoning in logic and Bayesian inference, if one accepts probabilities are a correct measure for degrees of belief towards a given statement. In a Bayesian setting it is possible to formulate any problem and incorporate into this formulation all the available knowledge about it, before any data is measured. In fact, as argued in the book by E.T. Jaynes [83], that is exactly what one should do when drawing conclusions about any parameter of a physical theory, given that always we would deal with an incomplete representation of the parameter space simply due to the finite data sample. One might be worried that this prior knowledge can ultimately bias the result, however it can be argued that if one phrases the prior information using the principle of maximal entropy coupled with reparameterization invariance (see Chapter 12 in [84], also see [85]), the prior will be automatically maximally ignorant about any particular parameter of our model. This is enough motivation for us to adopt the Bayesian inference as the main statistical inference machinery of this thesis.

The goal of any cosmological inference is to deduce what are the preferred values of a given set of model parameters Θ by the data D . In the Bayesian reasoning language, this means that we are interested in inferring the *posterior* probability distribution of a given model \mathcal{M} , $\mathcal{P}(\Theta|D, \mathcal{M}, I)$. The symbol I stands for any other background information about the problem we might have. The Bayesian approach then tells us we can use probability calculus to do reasoning, and hence we can apply the rule of conditional probabilities

$$\mathcal{P}(\Theta|D, \mathcal{M}, I) = \frac{\mathcal{P}(D|\Theta, \mathcal{M}, I)\mathcal{P}(\Theta|I)}{\mathcal{P}(D|\mathcal{M}, I)}. \quad (2.88)$$

The equation is otherwise known as the Bayes theorem. This equation relates the *likelihood*, $\mathcal{P}(D|\Theta, \mathcal{M}, I)$, of observing a particular dataset realization D , given the set of parameters Θ of model \mathcal{M} , supplemented by background information I . This is additionally weighted by the prior knowledge we have on the problem, which model \mathcal{M} conveys and it is formulated before receiving the data and having information I in addition at our disposal, $\mathcal{P}(\Theta|\mathcal{M}, I)$. The normalization factor, $\mathcal{P}(D|\mathcal{M}, I)$, denotes the Bayesian evidence and represents the model's preference for the given dataset. However, usually one is not interested in this quantity but rather in the preference of the data for a given model \mathcal{M} , i.e. $\mathcal{P}(\mathcal{M}|D, I)$. For example in Chapter 4 we will be interested in which spectral tilt for the WIMP dark matter decay does the data prefer better. Again, one can utilize the Bayes theorem and write

$$\mathcal{P}(\mathcal{M}|D, I) = \frac{\mathcal{P}(D|\mathcal{M}, I)\mathcal{P}(\mathcal{M}|I)}{\mathcal{P}(D|I)}, \quad (2.89)$$

where $\mathcal{P}(\mathcal{M}|I)$ denotes the prior belief in a given model \mathcal{M} , as dictated by background knowledge I , while $\mathcal{P}(D|I)$ denotes our prior belief in the plausibility of a given data

realization. However, one never calculates the posterior $\mathcal{P}(\mathcal{M}|D, I)$, since it only makes sense to compare two different models \mathcal{M}_1 and \mathcal{M}_2 and see which one of them is preferred by the data. Therefore, one is interested in calculating the following ratio

$$\frac{\mathcal{P}(\mathcal{M}_1|D, I)}{\mathcal{P}(\mathcal{M}_2|D, I)} = \frac{\mathcal{P}(D|\mathcal{M}_1, I)\mathcal{P}(\mathcal{M}_1|I)}{\mathcal{P}(D|\mathcal{M}_2, I)\mathcal{P}(\mathcal{M}_2|I)}. \quad (2.90)$$

This ratio is otherwise known as the *Bayesian odds factor* and is tightly related to the Occam's razor principle (see also [86]) as we demonstrate below.

A priori, the background knowledge I can be uninformative about the model choice and hence assign them equal probabilities $\mathcal{P}(\mathcal{M}_1|I) = \mathcal{P}(\mathcal{M}_2|I)$. The problem then reduces to calculating the Bayesian evidence ratio for the two models. It then follows that

$$\begin{aligned} \mathcal{P}(D|\mathcal{M}, I) &= \int \mathcal{D}\Theta \mathcal{P}(D|\Theta, \mathcal{M}, I)\mathcal{P}(\Theta|\mathcal{M}, I) \\ &\approx \int \mathcal{D}\Theta \exp \left[\ln \mathcal{P}(D|\hat{\Theta}, \mathcal{M}, I) + \frac{1}{2} (\Theta - \hat{\Theta})^i \mathcal{C}_{ij} (\Theta - \hat{\Theta})^j \right] \\ &= \mathcal{P}(D|\hat{\Theta}, \mathcal{M}, I) \sqrt{\frac{(2\pi)^{\mathcal{N}_\Theta}}{\det(\mathcal{C})}}, \end{aligned} \quad (2.91)$$

where we have assumed likelihood has a local maximum value $\hat{\Theta}$ and doesn't have very heavy tails, so that the Gaussian approximation holds to a good degree. Expanding around it then defines the curvature matrix

$$\mathcal{C}_{ij} \stackrel{\text{def}}{=} \left. \frac{\partial^2}{\partial \Theta^i \partial \Theta^j} (\ln \mathcal{P}(D|\Theta, \mathcal{M}, I)) \right|_{\Theta=\hat{\Theta}}.$$

The final step involves a simple Gaussian integral, where \mathcal{N}_Θ denotes the dimensionality of the parameter space. If the problem is such that the likelihood is a convex function of Θ , then there will exist a unique $\hat{\Theta}$ which will correspond to a global maximum. Otherwise, the only way of correctly estimating the integral is through sampling techniques, some of which will be covered in next subsections. In the limit of informative data, i.e. $\mathcal{N}_\Theta \ll \mathcal{N}_D$, with \mathcal{N}_D being the number of independent measurements, we have to leading order

$$\ln \det \left(\frac{1}{\mathcal{N}_D} \mathcal{C} \right) = \mathcal{N}_\Theta \ln \mathcal{N}_D + \det \left(\frac{1}{\mathcal{N}_\Theta} \mathcal{C} \right) \approx \mathcal{N}_\Theta \ln \mathcal{N}_D. \quad (2.92)$$

Therefore, applying this identity and taking the ln of Eq. (2.91) gives

$$\ln \mathcal{P}(D|\mathcal{M}, I) \approx \ln \mathcal{P}(D|\mathcal{M}, \hat{\Theta}, I) - \frac{\mathcal{N}_\Theta}{2} \ln \mathcal{N}_D. \quad (2.93)$$

As we can see, Bayesian evidence implicitly penalizes models with higher number of degrees of freedom, while it rewards the models which are better at explaining the given

data realization. Therefore, it is the balance of model complexity and model accuracy that determines the preference of the data. The Eq. (2.93) is also known as the *Bayesian information criterion* (BIC) and is used in Chapter 4 for distinguishing between different plausible explanations of the decay signal detected within the γ -ray data.

As we have seen, the Bayesian model selection involves calculating potentially high dimensional integrals. However, even when obtaining the posterior distribution one might be interested in integrating out uninteresting degrees of freedom of the model, to arrive at useful constraints for the parameters of interest. For example, one might be interested in marginalizing over the bias parameters introduced in Section 2.4 to obtain the posterior for cosmological parameters preferred by the data

$$\mathcal{P}(\Theta|\delta_g) = \frac{1}{\mathcal{P}(\delta_g)} \int \mathcal{D}\{b_O\} \mathcal{P}(\delta_g|\Theta, \{b_O\}) \mathcal{P}(\{b_O\}, \Theta)$$

where δ_g denotes the measured galaxy density. Depending on the form of the likelihood $\mathcal{P}(\delta_g|\Theta, \{b_O\})$ and the dimensionality of the bias parameter space, the above integral might be too difficult to perform analytically. In this case, we need to be satisfied with only approximating the posterior distribution, usually by calculating the estimators for its mean and covariance. For example, the unbiased estimator of the mean is given by

$$\langle \Theta \rangle_{\mathcal{P}(\Theta|\delta_g)} = \int_{\Theta} \mathcal{D}\Theta \Theta \mathcal{P}(\Theta|\delta_g) \Theta \approx \frac{1}{N} \sum_n \Theta_n,$$

with $\Theta_n \leftarrow \mathcal{P}(\Theta|\delta_g)$. (2.94)

Therefore, it is necessary to develop techniques capable of sampling from the intractable posterior distribution. In the next section, we present such a method, which is utilized throughout the thesis.

2.5.1 Hamiltonian Monte Carlo

In Chapter 3 we will be dealing with very high dimensional problems, briefly outlined in Section 2.6. This is a common feature of all field-level forward modelling approaches. For such problems, the conventional Monte Carlo sampling techniques are not very efficient. The reason is that the cost for generating an independent sample from a distribution with N degrees of freedom goes as $\sim N$ for a usual MCMC technique. This is a general feature of all random walk algorithms such as Metropolis-Hastings. A particularly efficient implementation of MH algorithm is available in [87]. Therefore, when the dimensionality is of order $\sim 10^6$ as we will see in Section 2.6 is the order of magnitude for field-level forward models, the sampling will be too inefficient. The Hamiltonian Monte Carlo algorithm [88] is specifically designed to tackle this problem. It achieves this feat in the following way

Each model parameter Θ^i , is associated with a corresponding conjugate momentum variable P^i which allows for defining the joint Hamiltonian as

$$\mathcal{H} = \frac{1}{2} P^i (M^{-1})_{ij} P^j - \ln \mathcal{L}(\Theta), \quad (2.95)$$

with \mathcal{L} being the model likelihood, which is a function of model parameters Θ^g , and M_{ij} representing the model's mass matrix. The role of the mass matrix is to approximate the local curvature of the parameter space as close as possible, for the most efficient exploration of the posterior [89]. The sample generation proceeds by solving the set of Hamiltonian equations

$$\begin{aligned} \frac{d\Theta^i}{du} &= \frac{\partial \mathcal{H}}{\partial P^i}, \\ \frac{dP^i}{du} &= -\frac{\partial \mathcal{H}}{\partial \Theta^i} \equiv \frac{\partial \mathcal{L}}{\partial \Theta^i}, \end{aligned}$$

with u parameterizing the Hamiltonian trajectory. When far away from the typical set, this formulation allows for driving the trajectories towards regions of high likelihood. Once the trajectories converge to the typical set, due to the conservation of Hamiltonian, the trajectories stay within the typical set and hence the obtained samples are guaranteed to also come from this set. Therefore, it is clear that the target distribution of the HMC chain is

$$\pi(P, \Theta) \sim e^{-\mathcal{H}(P, \Theta)} = \mathcal{L}(\Theta) e^{\frac{1}{2} P^T M^{-1} P}. \quad (2.96)$$

The reversibility and phase-space volume preservation of Hamiltonian dynamics allow for defining the following transition probability [88]

$$T(\Theta', P' | \Theta, P) = \min \left[1, e^{\mathcal{H}(P, \Theta) - \mathcal{H}(P', \Theta')} \right]. \quad (2.97)$$

As one can notice, if we end up in the typical set, due to the conservation of the Hamiltonian, we will keep on sampling with high acceptance rate from the target distribution. However, usually, the trajectory integration is implemented through discrete step size symplectic integrators, such as leap frog [90]. It is important the utilized integrators are symplectic, both due to approximate reversibility and volume preservation of such integrators.

In order to generate new samples with HMC, it is only necessary to sample the conjugate momenta from the zero centered Gaussian and covariance given by the mass matrix M . Resampling the momenta at each step guarantees the Markov property is preserved along the chain. This algorithm however comes with a cost of tuning properly the length of integration along the Hamiltonian trajectory, as well as tuning the mass

⁹Note here that the likelihood, \mathcal{L} , is a function of the model parameters, but it is a probability density in data values

Algorithm 1: Slice sampling

Data: Previous sample – Θ^* **Result:** New sample – Θ'

```

1
2  $\mathcal{L}^* \leftarrow \mathcal{L}(\Theta^*)$ 
3  $\ell \leftarrow (0, \mathcal{L}(\Theta^*))$ 
4
5  $r \leftarrow \mathcal{U}(0, 1)$ 
6  $\Theta_l \leftarrow \Theta^* - r\theta$ 
7  $\Theta_r \leftarrow \Theta^* + (1 - r)\theta$ 
8
9 while  $\mathcal{L}(\Theta_l) > \ell$  do
10 |  $\Theta_l \leftarrow \Theta_l - \theta$ 
11 end
12 while  $\mathcal{L}(\Theta_r) > \ell$  do
13 |  $\Theta_r \leftarrow \Theta_r + \theta$ 
14 end
15
16 while  $\mathcal{L}(\Theta') < \ell$  do
17 |  $\Theta' \leftarrow \mathcal{U}(\Theta_l, \Theta_r)$ 
18 | if  $\mathcal{L}(\Theta') > \ell$  then
19 | | return  $\Theta'$ 
20 | else
21 | | if  $\Theta' > \Theta^*$  then
22 | | |  $\Theta_r \leftarrow \Theta'$ 
23 | | else
24 | | |  $\Theta_l \leftarrow \Theta'$ 
25 | | end
26 | end
27 end

```

matrix. However, if tuned properly, it can achieve a very efficient scaling with dimensionality $\sim N^{1/4}$ [88]. Therefore, it is the algorithm of choice for the forward models presented in Section 2.6.

2.5.2 Slice sampling

A particularly interesting type of Markov chain Monte Carlo sampling (MCMC) technique is embodied in the slice sampling algorithm [91] (see also Section 29.7 in [86]).

Similarly, as for other Monte Carlo techniques, it can be utilized whenever one is capable of evaluating the value of the probability function at any given point in parameter space Θ^* . The main advantage of using this algorithm is its robustness, i.e. it doesn't require as much tuning as the HMC algorithm does, but has the advantage over standard Metropolis-Hastings algorithm for example in not requiring any tuning for the sampling interval. It is quite a simple algorithm, embodied in the pseudocode from Algorithm 1. A few lines require some further clarification. Lines 2-7 set the stage for the algorithm. As input, it only requires the initial sample Θ^* , which can for example be drawn from the prior $\mathcal{P}(\Theta)$, or be one of the previously obtained samples in the Markov chain. Lines 9-14 set up the interval (Θ_l, Θ_r) such that the adjacent points outside this interval have a likelihood value below ℓ . Note that this value $\ell \leftarrow \mathcal{U}(0, \mathcal{L}(\Theta^*))$ (line 3). This allows for preserving the Markov property of the chain, given that at each new step the state is completely randomized through drawing ℓ . Then in lines 16 - 27, a new sample is sought after within the interval $\Theta' \leftarrow (\Theta_l, \Theta_r)$. After a sample with a larger likelihood value is found it is accepted and returned, otherwise the interval is reset according to lines between 20 - 25.

The slice sampling algorithm is particularly efficient in exploring the 1D probability distributions, no matter if the distribution is multi-modal or not. Therefore, it is particularly useful when coupled to the HMC sampling block, for example for exploring low dimensional subspaces of the posterior, marginalizing out the parameters of no immediate interest (such as bias parameters for example). This is exactly the scheme used within the BORG framework (Section 2.6.1) as well as the EFTofLSS forward modelling framework from Chapter 3. Next, we focus on the field-level approaches to modelling the large-scale structure and explain different pieces of the models in more detail.

2.6 The field-level approach to structure formation

The EFTofLSS approach has so far been applied on the level of n -point statistics (see [14, 15] for example), however the question remains whether an approach modelling the full density field of galaxies could extract more useful information on cosmology than the approach utilizing summary statistics. These approaches are collectively labelled as *field-level* approaches. So far, most of the field-level studies have focused on the use of N-body simulations for solving the dark matter clustering problem with a simplified bias model on top to relate the underlying dark matter density field to the observed galaxy data [20, 23, 92–94]. However, different approaches utilizing deep neural nets and emulators are also emerging [25, 26, 95]. Nonetheless, which approach is chosen, all of them can be formulated in a Bayesian language (see Section 2.5), as trying to solve the following problem

$$\mathcal{P}(\Theta|\delta_g; \mathcal{M}) \sim \int \mathcal{D}\{\lambda_n\} \int \mathcal{D} \hat{s} \mathcal{P}(\delta_g|\Theta, \{\lambda_n\}, \hat{s}; \mathcal{M}) \mathcal{P}(\Theta, \{\lambda_n\}, \hat{s}|\mathcal{M}), \quad (2.98)$$

where $\Theta = \{\{\Omega_i\}, \sigma_8, \mathcal{H}, \dots\}$ represents the set of cosmological parameters of the model \mathcal{M} of interest, the δ_g represents the data such as galaxy density field for example, while the $\{\lambda_n\}$ and \hat{s} represent the set of nuisance parameters of the model \mathcal{M} and the initial conditions model \mathcal{M} requires. The field δ_g is typically represented on a Cartesian or Fourier space grid, where the physical size of the grid and the number of cells it has determines its Nyquist frequency, the maximal resolvable Fourier mode of the grid. The Nyquist frequency provides a natural cutoff scale for any bias expansion used in describing the tracer field. This becomes especially important for the EFT field-level forward models, as described in Chapter 3.

The proportionality sign in Eq. (2.98) emphasizes that we're not interested in the overall normalization of the equation, $\mathcal{P}(\delta_g; \mathcal{M})$, at this stage, which will be necessary only at the model comparison stage (see Section 2.5). Given that all the work in this thesis is focused on either the EFTofLSS forward modelling approach (Chapter 3) or the N-body approach (Chapters 5 and 4) in the following subsections a short overview of the two forward modelling approaches is given.

2.6.1 BORG forward model

The Bayesian Origin Reconstruction in Galaxies (BORG) forward model has been first outlined in [96] and subsequently improved in [19–22] with ever present ongoing development. It is surprisingly simple to formulate in terms of Eq. (2.98)

$$\begin{aligned} \mathcal{P}(\Theta|N_g, \mathcal{M}_1) &= \int \mathcal{D}\{\lambda_n\} \int \mathcal{D}\hat{s} \mathcal{P}(\delta_g|\Theta, \{\lambda_n\}, \hat{s}; \mathcal{M}_1) \mathcal{P}(\{\lambda_n\}, \Theta, \hat{s}|\mathcal{M}_1) \\ &= \int \mathcal{D}\{\lambda_n\} \int \mathcal{D}\hat{s} \mathcal{P}(N_g|\delta_{\text{det}}; \mathcal{M}_1) \mathcal{P}(\delta_{\text{det}}|\Theta, \{\lambda_n\}, \hat{s}; \mathcal{M}_1) \mathcal{P}(\{\lambda_n\}, \Theta, \hat{s}|\mathcal{M}_1) \end{aligned} \quad (2.99)$$

where the data is taken to be the observed galaxy counts N_g . The individual terms are given as

$$\begin{aligned} \mathcal{P}(N_g|\delta_{\text{det}}; \mathcal{M}_1) &= \prod_{\mathbf{x}_i} \frac{(\delta_{\text{det}}^{\mathbf{x}_i})^{(N_g)_{\mathbf{x}_i}} e^{-\delta_{\text{det}}^{\mathbf{x}_i}}}{((N_g)_{\mathbf{x}_i})!}, \\ \mathcal{P}(\delta_{\text{det}}|\Theta, \{\lambda_n\}, \hat{s}; \mathcal{M}_1) &= \prod_{\mathbf{x}_i} \delta_D(\delta_{\text{det}}^{\mathbf{x}_i} - D^{\mathbf{x}_i}(\hat{s}, \{\lambda_i\}, \Theta)). \end{aligned} \quad (2.100)$$

The first line of the above equation describes the probability of observing the galaxy counts N_g inside the observed volume, given the expected value of the deterministic model prediction for the density field for the same volume, δ_{det} . It is given simply through the Poisson likelihood, with the product running over each Cartesian grid element, centered around \mathbf{x}_i . The second line describes the probability of realizing the $\delta_{\text{det}}^{\mathbf{x}_i}$ at the position

\mathbf{x}_i within the volume, given the deterministic mapping, D , from the initial conditions of the model, \hat{s} . This mapping depends on the given set of cosmological parameters Θ , the set of nuisance parameters $\{\lambda_i\}$ and the model's initial conditions \hat{s} . The mapping D can further be decomposed into the following set of mappings

$$D = \mathcal{R} \circ \mathcal{S}(\Theta) \circ \mathcal{B}(\{\lambda_i\}) \circ G(\Theta, \delta_i(\hat{s}, \Theta)) \quad (2.101)$$

where the \mathcal{R} encompasses all survey mask effects, which can be quite difficult to handle at the level of summary statistics (e.g. [13]), the \mathcal{B} represents the bias model, the \mathcal{S} the mapping to redshift space, which in general depends on the assumed cosmology (see for example [97, 98]) and finally the G represents the deterministic gravitational evolution, mapping the initial overdensities δ_i into the final dark matter density field. The initial overdensity field δ_i is a deterministic function of initial conditions of the forward model \hat{s} as will be elaborated further below. For the operator G , there are multiple choices within the BORG framework one of them being a differentiable N-body simulation implemented through Particle Mesh model (see [23]), which was also used for the results of Chapters 5 and 4. Depiction of all these forward modelling elements is available in Fig. 5.1. The bias model \mathcal{B} is usually chosen to follow the semi-empirical relation from [99] which contains 4 degrees of freedom, which corresponds to a model similar to the LIMD bias expansion from Eq. (2.67).

As a final step, one needs to also evaluate the integrals from Eq. (2.99). This is done through the use of Monte Carlo methods such as block sampling. The blocks are split into the slice sampling block for λ_n and Θ parameters, and the Hamiltonian Monte Carlo sampling block for sampling the initial conditions \hat{s} (see Section 2.5). In order to perform this step, one also needs to specify the prior distributions for Θ , $\{\lambda_n\}$ and \hat{s} . The priors for Θ , $\{\lambda_n\}$ are usually taken to be drawn from a uniform distribution, for example taking reasonable bounds informed by the latest Planck results [5], while the initial conditions are sampled from a white Gaussian field, i.e. $\hat{s}^{\mathbf{x}_i} \leftrightarrow \mathcal{G}(0, \mathbf{1})$. The choice of the white Gaussian random field makes the setting of initial overdensities, δ_i , needed by the mapping G generic. The initial overdensities, δ_i , for the Particle Mesh can be initialized through $\delta_i(\mathbf{k}; \Theta) = \sqrt{P_L(k; \Theta)} \hat{s}(\mathbf{k})$. To see that δ_i indeed has the correct correlation structure one can look at

$$\begin{aligned} \langle \delta_i(\mathbf{k}; \Theta) \delta_i(-\mathbf{k}; \Theta) \rangle' &= \langle \sqrt{P_L(k; \Theta)} \hat{s}(\mathbf{k}) \sqrt{P_L(k; \Theta)} \hat{s}(-\mathbf{k}) \rangle' = P_L(k; \Theta) \underbrace{\langle \hat{s}(\mathbf{k}) \hat{s}(-\mathbf{k}) \rangle'}_1, \\ &= P_L(k; \Theta). \end{aligned} \quad (2.102)$$

Therefore in total, the BORG forward model has on the order of $\sim 10^6 - 10^7$ free parameters, depending on the resolution of the initial conditions grid. This high dimensional posterior is a daunting inference task and is a shared feature among field-level forward modelling approaches. It is necessary to have a robust forward model in order to efficiently explore such a high dimensional space.

Even with all the detailed modelling of the BORG framework, questions still remain on whether the likelihood from Eq. (2.100) is well motivated and whether the bias model is indeed applicable to the range of scales considered. As we have seen in Section 2.4, the choice of scale up to which one strives to describe the tracer field implicitly defines the nature of the bias expansion to be used. Given that the EFTofLSS approach to the clustering problem offers a natural formulation of both the bias expansion and the nature of the underlying stochastic field, it is a worthwhile effort to try and join the field-level approach presented in this chapter with the EFTofLSS framework.

2.6.2 EFTofLSS field-level forward model

One of the issues raised in the previous section is whether the Poisson likelihood is indeed a good choice for modelling the observed halo distribution. This is tightly related to the assumption on the noise distribution. Namely, the most general forward model with additive noise can be written as

$$\delta_d(\mathbf{x}) = \delta_{\text{det}}[\Theta, \{b_O\}, \hat{s}](\mathbf{x}) + \epsilon(\mathbf{x}; \Sigma(\mathbf{x})), \quad (2.103)$$

where once again Θ represents all cosmological parameters of the model, Σ represents the correlation structure of the noise, the \hat{s} initial conditions of the model and $\{b_O\}$ the set of bias parameters the model implements such that

$$\delta_{\text{det}}[\Theta, \{b_O\}, \hat{s}](\mathbf{x}) = \sum_O b_O [O][\Theta, \hat{s}](\mathbf{x}), \quad (2.104)$$

with $[O] \in \{\delta, [\delta]^2, (\partial_i \partial_j \Psi_\Lambda)^2, \dots\}$. This bias expansion arises as a direct consequence of general considerations from Section 2.4 and therefore is motivated by the fundamental principles of the theory, unlike the semi-empirical relation used in Section 2.6.1. The Bayesian reasoning suggests that the parameter joint posterior is given by

$$\begin{aligned} \mathcal{P}(\Theta, \{b_O\}, \hat{s} | \delta_d) &\sim \int \mathcal{D}\epsilon \mathcal{P}(\delta_d | \Theta, \{b_O\}, \hat{s}, \epsilon) \mathcal{P}(\Theta, \{b_O\}, \hat{s}) \mathcal{P}_\epsilon(\epsilon; \Sigma) \\ &= \int \mathcal{D}\epsilon \delta_D(\delta_d - \delta_{\text{det}}[\Theta, \{b_O\}, \hat{s}] - \epsilon) \mathcal{P}(\Theta, \{b_O\}) \mathcal{P}_\epsilon(\epsilon; \Sigma) \\ &= \mathcal{P}_\epsilon(\delta_d - \delta_{\text{det}}[\Theta, \{b_O\}, \hat{s}]; \Sigma) \mathcal{P}(\Theta, \{b_O\}, \hat{s}), \end{aligned} \quad (2.105)$$

neglecting the normalization term. In the above equation, \mathcal{P}_ϵ represents the noise probability density function, and as can be seen, the obtained likelihood will follow the same distribution law. This is a direct consequence of assuming additive noise.

In case of large-scale structure, additive noise assumption is completely consistent and was already suggested to us by the analysis of terms appearing within the effective stress tensor τ^{ij} . The leading order counter terms from Eq. (2.48), directly support

this assumption. What we obtained in Section 2.4 is that the leading order correction to the shot noise field of biased tracers, comes with a correlation structure scaling as $\sim k^2$ in Fourier space. Therefore, the EFTofLSS suggests that corrections to the Poisson assumption for the noise exist and should be included already at leading order. Therefore, this renders the assumption of the Poisson likelihood from the previous section as too restricting. For example, one could also ask of why the log-normal distribution would not be a better choice (see e.g. [100, 101]).

On the other hand, the EFTofLSS suggests clearly the form of all higher order stochastic counter terms (see Section 2.8 in [31]) that should be included, and gives a clear prescription at which order they should appear. This motivates the form of the likelihood itself directly from the theory.

The only remaining condition for the EFTofLSS forward modelling approach is that it should capture the full spectrum of the tracer n -point statistics. The BORG forward model naturally captures all n -point statistics of the matter field through simply forward evolving the field fully non-linearly utilizing the N-body simulation. However, it is not clear whether this is also true for the underlying *tracer* field, like the galaxy clustering field, given the simple bias assumption usually applied [99]. In contrast, as we have seen, the effective field theory approach relies on a perturbation series expansion to model the final tracer field realization where the terms are ordered according to the perturbation theory and are ultimately controlled by the requested modelling precision. In the following section, we demonstrate that the EFTofLSS forward model presented here is indeed capable of capturing higher order statistics of the given tracer field generically, thus motivating the pursuit for an EFTofLSS field-level forward modelling approach undertaken in Chapter 3.

From n -point to field-level

At the field-level, besides noise and bias parameters, one also needs to specify the set of initial conditions \hat{s} , in order to generate a forward model realization. Given that all these degrees of freedom are of secondary interest, it is necessary to marginalize over them in order to obtain our constraints on cosmology. In other words

$$\mathcal{P}(\Theta|\delta_g) \sim \left[\int \mathcal{D}\{b_O\} \mathcal{P}(\{b_O\}) \int \mathcal{D}\hat{s} \mathcal{P}(\hat{s}) \int \mathcal{D}\epsilon \mathcal{P}(\epsilon) \mathcal{P}(\delta_g|\Theta, \{b_O\}, \hat{s}, \epsilon) \right] \mathcal{P}(\Theta), \quad (2.106)$$

again dropping the normalization term. The likelihood term can be rewritten in the following way

$$\begin{aligned} \mathcal{P}(\delta_g|\Theta) &= \int \mathcal{D}\{b_O\} \mathcal{P}(\{b_O\}) \int \mathcal{D}\hat{s} \mathcal{P}(\hat{s}) \int \mathcal{D}\epsilon \mathcal{P}(\epsilon) \delta_D(\delta_d - \delta_{\text{det}}[\Theta, \{b_O\}, \hat{s}] - \epsilon) \\ &= \int \mathcal{D}\{b_O\} \mathcal{P}(\{b_O\}) \langle \delta_D(\delta_d - \delta_{\text{det}}[\Theta, \{b_O\}, \hat{s}] - \epsilon) \rangle, \end{aligned} \quad (2.107)$$

using the expression for the forward model from Eq. (2.103). This expectation value of the Dirac delta functional, δ_D , can be cast into the following integral form, using the Fourier representation of the Dirac delta (dropping the arguments for clarity)

$$\int \mathcal{D} X \left\langle e^{\int_{\mathbf{q}} iX(\delta_g - \delta_{\text{det}} - \epsilon)} \right\rangle = \int \mathcal{D} X e^{-\int_{\mathbf{q}} iX \delta_g} \left\langle e^{\int_{\mathbf{q}} iX(\delta_{\text{det}} + \epsilon)} \right\rangle. \quad (2.108)$$

As already discussed, the EFTofLSS gives a clear prescription for the correlation structure of the noise field. At leading order, it is diagonal in Fourier space, and is analytic in even powers of Fourier modes k^{2n} (see Section 2.4 as well as Section 3.2.3). This then motivates a zero mean Gaussian prior with correlation structure $\Sigma(k)$ for the noise field. Therefore, one can readily perform the integral over the $\mathcal{P}(\epsilon)$

$$\begin{aligned} \left\langle e^{\int_{\mathbf{q}} iX(\delta_{\text{det}} + \epsilon)} \right\rangle &= \int \mathcal{D} \hat{s} \hat{\mathcal{P}}(\hat{s}) e^{\int_{\mathbf{q}} iX \delta_{\text{det}}} \int \mathcal{D} \epsilon \mathcal{P}(\epsilon) e^{\int_{\mathbf{q}} iX \epsilon} \\ &= \int \mathcal{D} \hat{s} \hat{\mathcal{P}}(\hat{s}) e^{\int_{\mathbf{q}} iX \delta_{\text{det}}} \int \mathcal{D} \epsilon e^{\int_{\mathbf{q}} (iX \epsilon - \frac{1}{2} \epsilon^2(\mathbf{q}) \Sigma^{-1}(\mathbf{q}))} \\ &= e^{\frac{1}{2}((iX)^2(\mathbf{q}) \Sigma(\mathbf{q}))} \int \mathcal{D} \hat{s} \hat{\mathcal{P}}(\hat{s}) e^{\int_{\mathbf{q}} iX \delta_{\text{det}}}, \end{aligned} \quad (2.109)$$

where we perform a simple Gaussian integration step from second to third line. The remaining expectation value can be rewritten in a more enlightening form

$$\left\langle e^{\int_{\mathbf{q}} iX \delta_{\text{det}}} \right\rangle_{\mathcal{P}(\hat{s})} = \sum_n \int_{\mathbf{q}_1} iX(\mathbf{q}_1) \cdots \int_{\mathbf{q}_n} iX(\mathbf{q}_n) \langle \delta_{\text{det}}(\mathbf{q}_1) \cdots \delta_{\text{det}}(\mathbf{q}_n) \rangle_{\mathcal{P}(\hat{s})} \equiv Z_{\hat{s}}[iX], \quad (2.110)$$

which is nothing else than an expression for a generating functional with the current being given by iX . Expanding further the δ_{det} term using the general bias expansion from Section 2.4 shows that the forward model of this form can systematically generate all n -point functions of the resulting tracer field beyond the leading order. In fact reference [102] shows that the field-level EFT likelihood derived above in combination with 2nd order Lagrangian perturbation theory (see Appendix B), is capable of capturing the cosmological information contained in leading- and next-to-leading order power spectrum and leading order bispectrum. Furthermore, reference [103] also shows that the description of curvature and tidal effects at the typical scale of BAO^{10} is also captured with the field-level EFT forward model which effectively contains the information on the 4-point and higher-order statistics. Therefore, putting all together back into Eq. (2.108) and integrating over the source field X , gives (reintroducing the explicit \hat{s} dependence for clarity)

$$\left\langle \int \mathcal{D} X e^{\int_{\mathbf{q}} (iX \delta_{\text{det}}[\hat{s}])(\mathbf{q}) + \frac{1}{2}((iX)^2(\mathbf{q}) \Sigma(\mathbf{q}))} \right\rangle_{\mathcal{P}(\hat{s})} = |\Sigma|^{-1/2} \left\langle e^{-\frac{1}{2} \int_{\mathbf{q}} (\delta_a(\mathbf{q}) - \delta_{\text{det}}[\hat{s]}(\mathbf{q}))^2 \Sigma^{-1}(\mathbf{q})} \right\rangle_{\mathcal{P}(\hat{s})}. \quad (2.111)$$

¹⁰Baryon acoustic oscillations

This demonstrates that at leading order, the field-level likelihood of the EFTofLSS approach takes the form of a Gaussian with mean given by the forward model prediction $\delta_{\text{det}}[\hat{s}]$ and diagonal covariance Σ . The final expression of the cosmological posterior will then be given as

$$\mathcal{P}(\Theta|\delta_g) = \left\langle |\Sigma|^{-1/2} e^{-\frac{1}{2} \int_{\mathbf{q}} (\delta_d(\mathbf{q}) - \delta_{\text{det}}[\{b_O\}, \hat{s}](\mathbf{q}))^2 \Sigma^{-1}(\mathbf{q})} \right\rangle_{\mathcal{P}(\{b_O\}, \hat{s})} \mathcal{P}(\Theta). \quad (2.112)$$

In fact, given that the bias parameters enter linearly into the bias expansion, in this case one can even analytically marginalize over the bias parameters and obtain the marginalized likelihood expression from Eq. (3.16) (see also [104, 105]). This likelihood expression has in turn more desirable properties, as demonstrated in Fig. 3.9.

Note that the same procedure can be repeated with a more general noise model, introducing the higher order stochastic terms such as $\sim \epsilon_\delta \delta$ (see Chapter 2.8 in [31]) which will then generate a more complicated covariance structure but keep the Gaussian form intact.

For more details on how the specific forward models are constructed utilizing this approach, reader is referred to Section 3.2 and the following references [30, 102, 104, 106, 107] as well as [97, 108]. The latter two references however use the residual power spectrum as their likelihood function and by this effectively average out the information contained in the modes falling in the given power spectrum bin. On the other hand, the former as well as the work presented in Chapter 3 extracts cosmological information at the level of individual modes.

CHAPTER 3

FIELD-LEVEL INFERENCE WITH THE EFT LIKELIHOOD

The following material was first presented in [109], with me as a first author and Fabian Schmidt, Minh Nguyen and Martin Reinecke as co-authors. Paper is submitted to Journal of Cosmology and Astroparticle Physics.

Abstract

Analyzing the clustering of galaxies at the field-level in principle promises access to all the cosmological information available. Given this incentive, in this chapter we investigate the performance of field-based forward modeling approach to galaxy clustering using the effective field theory (EFT) framework of large-scale structure (LSS). We do so by applying this formalism to a set of consistency and convergence tests on synthetic datasets. We explore the high-dimensional joint posterior of LSS initial conditions by combining Hamiltonian Monte Carlo sampling for the field of initial conditions, and slice sampling for cosmology and model parameters. We adopt the Lagrangian perturbation theory forward model from [30], up to second order, for the forward model of biased tracers. We specifically include model mis-specifications in our synthetic datasets within the EFT framework. We achieve this by generating synthetic data at a higher cutoff scale Λ_0 , which controls which Fourier modes enter the EFT likelihood evaluation, than the cutoff Λ used in the inference. In the presence of model mis-specifications, we find that the EFT framework still allows for robust, unbiased joint inference of a) cosmological parameters — specifically, the scaling amplitude of the initial conditions — b) the initial conditions themselves, and c) the bias and noise parameters. In addition, we show that in the purely linear case, where the posterior is analytically tractable, our samplers fully

explore the posterior surface. We also demonstrate convergence in the cases of nonlinear forward models. Our findings serve as a confirmation of the EFT field-based forward model framework developed in [102, 104–107, 110], and as another step towards field-level cosmological analyses of real galaxy surveys.

3.1 Introduction

Current and future galaxy surveys such as DESI [6], Euclid [7], PFS [8], and the Vera Rubin Observatory [10] offer a wealth of modes for probing the physics of structure formation. The traditional approach to cosmology inference from galaxy clustering is to compress the galaxy density field into summary statistics, such as two-point (see [11] and references therein), three-point [12–15], and four-point functions [16–18].

An alternative approach, and the one we follow in this chapter, attempts to extract information at the field-level, by explicitly forward modeling the entire observed galaxy density field including all the relevant physics and observational effect. This physical Bayesian forward modeling approach [20, 23, 92–94] in principle allows for exploiting information on cosmology beyond n -point functions via explicit marginalization over the initial conditions. While the amount of cosmological information available beyond the low-order n -point functions is still unclear, this approach at the very least allows for a consistent treatment of Baryon Acoustic Oscillation reconstruction [103, 104], and thus is well motivated. Observational systematic effects can be explicitly encoded into the forward model (e.g. [93, 111]), and might be easier to disentangle from cosmological signals at the field-level as compared to summary statistics.

So far, field-level inference approaches have typically used an empirical galaxy bias model and a simplified likelihood to infer from galaxy clustering data. Using dark matter halos in N -body simulations as the reference data, [112] demonstrated that most empirical bias models and likelihoods that have been widely adopted by this approach thus far (e.g. [23, 93]) can significantly bias the inferred cosmological fields. Therefore, the key issue for this approach currently lies in a rigorous physical model, or alternatively a sufficiently flexible effective model (e.g. [25, 26, 95]) to connect the matter and tracer fields.

The effective field theory of large-scale structure (EFTofLSS) [27–29] provides a systematic way of incorporating the complex nonlinear physics of galaxy formation on small scales, by using the fact that galaxy formation is spatially localized, and that galaxies and matter comove on large scales (the latter is ensured by the equivalence principle). In particular, the EFTofLSS provides a parametric model for the matter-tracer relation up to the given order in matter density perturbations for any tracer-field of interest (see [31] for a review). This in turn allows for robust extraction of cosmological information from the tracer data up to quasilinear scales [15, 113], i.e. wavenumbers smaller than the nonlinear scale wavenumber k_{NL} ($k_{\text{NL}} \sim 0.2 h \text{ Mpc}^{-1}$ at $z = 0$). Until recently, the EFT predictions were restricted to summary statistics, but Refs. [102, 106, 110] (see [108]

for related work) presented a derivation of a field-level, EFT-based forward model and likelihood. The work described in this chapter a follow up to the series of EFT likelihood papers [30, 102, 104, 106, 107], with the crucial addition of the *marginalization over the initial conditions field by explicitly sampling it*. Specifically, we test the consistency of the EFT likelihood – previously demonstrated for fixed initial conditions – on a set of synthetic tracer data and range of forward models including model mis-specification. The tests of the EFT likelihood approach presented here serve as a crucial stepping stone towards applying the method to more realistic tracers such as dark matter halo or simulated galaxy catalogs. Throughout this chapter, we use the following fiducial cosmology: $\Omega_m = 0.3, \Omega_\Lambda = 0.7, h = 0.7, n_s = 0.967, \sigma_8 = 0.85$ and a box size of $L = 2000 h^{-1} \text{Mpc}$. We keep cosmological parameters fixed to the fiducial ΛCDM cosmology, but vary a scaling parameter α multiplying the initial conditions, which corresponds to varying σ_8 while keeping all other cosmological parameters fixed.

The structure of the chapter is as follows. In section Section 3.2, we elaborate on the specific forward models we use for consistency tests of the EFT likelihood. We first describe the simple toy models LINEAR and then physical models based on Lagrangian perturbation theory (LPT) 1LPT, 2LPT in Section 3.2.1 and Section 3.2.2, respectively. Next, we give the description of the EFT likelihoods (Section 3.2.3) and the expression for the full field-level posterior being sampled in Section 3.3.1. We then outline the sampling methods and the code implementation in Section 3.3.2 and Section 3.3.3, respectively. Afterwards, we describe how we generate the synthetic datasets in Section 3.4. Finally, we discuss our results in Section 3.5. We conclude and discuss future outlook in Section 3.6. In the appendices, additional details complementing the main results in Section 3.5 are presented and discussed.

3.2 Forward models

In this section, we present all forward models used in Section 3.2.1 and Section 3.2.2. Next, we describe the EFT likelihoods, the last piece of the inference framework, in Section 3.2.3.

We begin by discussing the common properties of the forward models. All forward models we consider in this chapter aim at modeling the following quantity

$$\delta_d(\mathbf{x}) \equiv \frac{n_d(\mathbf{x})}{\bar{n}_d} - 1 = \delta_{\text{det}}(\mathbf{x}) + \epsilon(\mathbf{x}), \quad (3.1)$$

with n_d representing the number density field of synthetic tracers, \bar{n}_d its spatial mean, and δ_d the fractional overdensity. Throughout this chapter, we will work with tracers defined at a fixed time τ , corresponding to today's epoch in the fiducial cosmology. Hence, we will drop the time argument for clarity.

We effectively marginalize over \bar{n}_d by working with δ_d and excluding the $\vec{k} = 0$ mode from the analysis. Thus, the field-level forward models consist of two parts: the mean-field prediction $\delta_{\text{det}}(\mathbf{x})$, which is a deterministic function of the initial conditions, and the likelihood, constructed from the assumptions about the underlying noise field $\epsilon(\mathbf{x})$. Next, we describe the different deterministic forward models used, before turning to the likelihoods which couple to all of these deterministic forward models.

The general form of the deterministic contribution is

$$\delta_{\text{det}}(\mathbf{x}) = \sum_O b_O[O](\mathbf{x}), \quad (3.2)$$

where $[O]$ denote the renormalized bias operators and b_O the corresponding coefficients (see Section 2.4 for more details). Depending on the gravity model and the bias model we use, Eq. (3.2) takes on different forms. Since we utilize an EFT approach, our forward models are defined for a specific cutoff Λ . The motivation for this cutoff is twofold:

1. The EFT model developed applies only to large scales. Hence, we want to restrict the likelihood evaluation only to the modes below the cutoff $k < \Lambda$. This in turn means that we need to apply a sharp- k cut to the field operators $O \rightarrow O_\Lambda$.
2. As first pointed out by [28], and then shown in detail in [106], it is also necessary to perform a cutoff at the level of initial conditions $\Lambda_{\text{in}} \equiv \Lambda$. This allows for proper renormalization of the dynamical evolution of the large-scale modes we want to model.

Throughout, we generate the synthetic data at a higher or equal cutoff Λ_0 than the value Λ used in the inference, motivated by the fact that real-world tracers resemble data with $\Lambda_0 \gg \Lambda$. In fact, real data effectively has no cutoff, i.e. $\Lambda_0 \rightarrow \infty$. Below, we describe the specifics of the forward models employed throughout the chapter. We focus first on the simplest limiting cases, which involve only linear density fields. We then explain how we build up the 1LPT and 2LPT forward models.

3.2.1 LINEAR forward models

In LINEAR forward models, the gravity model is restricted to linear evolution, which is incorporated by applying the appropriate transfer function to the initial conditions. On top of this, we also include tracer bias. We consider two different bias expansions. The first one involves only the linear bias b_δ and can be expressed as follows

$$\delta_{\text{det},\Lambda}^{\text{LINEAR}_1}(\mathbf{k}) = b_\delta \delta_\Lambda^{(1)}(\mathbf{k}), \quad (3.3)$$

$$\begin{aligned} \delta_\Lambda^{(1)}(\mathbf{k}) &\equiv W_\Lambda(k) \delta^{(1)}(\mathbf{k}) \\ &= W_\Lambda(k) \alpha T(k) \hat{s}(\mathbf{k}), \end{aligned} \quad (3.4)$$

where $\hat{s}(\mathbf{k})$ denotes a unit Gaussian field which describes the initial conditions (see also discussion in Section 3.3.1), while $T(k)$ denotes the transfer function (recall that we keep the time fixed and implicit throughout). Since we keep the cosmological parameters fixed, we do not write them explicitly here. The scaling parameter α is defined such that, for $\alpha = 1$, $\delta^{(1)}(\mathbf{k})$ corresponds to a realization of the linear density field in the fiducial cosmology. $W_\Lambda(k)$ denotes the isotropic sharp- k filter¹ at $k = \Lambda$. For this forward model (Eq. (3.3)), it is possible to analytically derive the posterior for the initial conditions \hat{s} (see Appendix 3.8.1 for detailed calculation), and hence to test whether our inference approach fully explores the posterior in this case.

The simplest extension of the above is to include also the quadratic field with the corresponding b_{δ^2} bias parameter:

$$\delta_{\text{det},\Lambda}^{\text{LINEAR}2}(\mathbf{k}) = b_\delta \delta_\Lambda^{(1)}(\mathbf{k}) + b_{\delta^2} W_\Lambda(k) \int_{\mathbf{k}'} \delta_\Lambda^{(1)}(\mathbf{k} - \mathbf{k}') \delta_\Lambda^{(1)}(\mathbf{k}'), \quad (3.5)$$

with $\delta_\Lambda^{(1)}$ given by Eq. (3.3). The bias expansion here is not complete in the EFT sense; we employ this forward model merely as the simplest possible generalization from an entirely linear forward model. The complete second-order bias expansion is considered in the next section. Nevertheless, due to the nonlinearity induced by the quadratic bias, the \hat{s} posterior is non-Gaussian and it is non-trivial to make exact statements about its statistical moments (see Appendix 3.9 for further discussion). The models given by Eqs. (3.3)–(3.5) thus serve as toy models for which there exists full or approximate analytical expression of the \hat{s} posterior. They both assume all the relevant information is contained within the linear density field and do not involve nontrivial gravitational displacements, which are however essential when attempting field-level inference on real data.

3.2.2 1LPT and 2LPT forward models

The forward models in Eqs. (3.3)–(3.5) are valid only for describing tracers at linear order. In order to obtain more accurate descriptions at higher orders – which properly account for gravitational evolution – we turn to Lagrangian perturbation theory. The Lagrangian formulation of structure formation captures the effect of bulk flows non-perturbatively. This is especially useful for forward modeling. To be more precise, we consider a first- and second-order LPT, labeling them with 1LPT and 2LPT, respectively. For a complementary review of LPT reader is referred to Appendix B. The following closely follows the more general approach undertaken in [30].

We begin by writing the Eulerian position along the fluid line at conformal time τ as

$$\mathbf{x}_\text{fl}(\mathbf{q}, \tau) = \mathbf{q} + \boldsymbol{\psi}(\mathbf{q}, \tau), \quad (3.6)$$

¹See Appendix 3.7 for more details and our Fourier space convention.

with \mathbf{q} being the Lagrangian coordinate, $\boldsymbol{\psi}$ the displacement field and $\lim_{\tau \rightarrow 0} \boldsymbol{\psi}(\mathbf{q}, \tau) = 0$. We also re-instate the explicit τ dependence throughout this overview for clarity. Combining the mass conservation condition and the Poisson equation for the non-relativistic (cold) dark matter fluid yields the following equation of motion for $\boldsymbol{\psi}$ [30, 114–116]

$$\text{tr} [(\mathbb{1} + \mathbf{M})^{-1}(\mathbf{M}'' + \mathcal{H} \mathbf{M}')] = -\frac{3}{2} \Omega_m \mathcal{H}^2 [|\mathbb{1} + \mathbf{M}|^{-1} - 1], \quad (3.7)$$

where $M_{ij}(\mathbf{q}, \tau) \equiv \partial_{(q_i} \psi_{j)}(\mathbf{q}, \tau)$ is the symmetric part of the Lagrangian distortion tensor, primes denote derivatives with respect to τ , Ω_m is the corresponding matter density parameter and \mathcal{H} the conformal Hubble rate. We restrict to the symmetric part of $\partial_{q_i} \psi_j$ throughout, since the antisymmetric part, corresponding to the curl of $\boldsymbol{\psi}$, is only nonzero at third order in perturbations, while here we restrict ourselves to second order. Thus, $\boldsymbol{\psi}$ is a longitudinal vector and can be written as

$$\boldsymbol{\psi}(\mathbf{q}, \tau) = \frac{\nabla}{\nabla^2} \sigma(\mathbf{q}, \tau); \quad \sigma(\mathbf{q}, \tau) \equiv \text{tr} \mathbf{M}(\mathbf{q}, \tau). \quad (3.8)$$

Lagrangian perturbation theory then proceeds by expanding [117]

$$\mathbf{M} = \mathbf{M}^{(1)} + \mathbf{M}^{(2)} + \dots, \quad (3.9)$$

and analogously for $\boldsymbol{\psi}$ and σ , where $\mathbf{M}^{(n)}$ involves exactly n powers of the linear density field $\delta^{(1)}$. In fact, $\sigma^{(1)}(\mathbf{q}, \tau) = -\delta^{(1)}(\mathbf{q}, \tau)$. For a given expansion history, Eq. (3.7) can be formally integrated to yield recurrence relations relating $\mathbf{M}^{(n)}$ to the lower-order contributions [114–116].

The perturbative contributions to \mathbf{M} in Eq. (3.9) can serve as building blocks for a general bias expansion of the form given in Eq. (3.2). The reason is that the Lagrangian distortion tensor along the fluid trajectory captures all leading gravitational observables for a comoving observer (see section 2.5 in [31] and [66]). Specifically, one needs to construct all scalar contractions of the $\mathbf{M}^{(i)}$ that are relevant at the given order n . At second order, this yields

$$O_L(\mathbf{q}, \tau) \in \left\{ \text{tr} [\mathbf{M}^{(1)}] (\mathbf{q}, \tau), (\text{tr} [\mathbf{M}^{(1)}] (\mathbf{q}, \tau))^2, \text{tr} [\mathbf{M}^{(1)} \mathbf{M}^{(1)}] (\mathbf{q}, \tau) \right\}, \quad (3.10)$$

where we emphasize that $O_L \equiv O(\mathbf{q}, \tau)$ is in Lagrangian coordinates. Thus, up to second order, we require three distinct bias operators and the associated bias coefficients.

In order to obtain the corresponding Eulerian fields, which is the space in which the data is obtained, we use a weighted particle approach [30, 108]. That is, we consider $(N_g^{\text{Eul}})^3$ particles on a regular grid in \mathbf{q} , and assign each of them 3 weights corresponding to the three bias operators. Then, each of these particles is displaced from \mathbf{q} to $\mathbf{x} = \mathbf{q} + \boldsymbol{\psi}(\mathbf{q}, \tau)$, and the masses are deposited to the grid using a mass-conserving assignment scheme (we choose cloud-in-cell assignment here). This yields the three Eulerian operators corresponding to the Lagrangian operators listed above. In fact, we

replace the weight field $\text{tr}[\mathbf{M}^{(1)}]$ with a unit weight field, so that the resulting Eulerian field is the LPT matter density field $\delta_{n\text{LPT}}$ ($n = 1, 2$), and use this linear bias term in Eulerian frame.

Finally, we also include the leading-order higher-derivative bias contribution (see sec 2.6 in [31]), $\nabla^2 \delta_{n\text{LPT}}$, computed in the Eulerian frame, in order to capture the finite spatial size of the regions forming the tracer of interest. Therefore, the full list of bias parameters and corresponding operators is

$$\begin{aligned} \delta_{\text{det},\Lambda}^{n\text{LPT}}(\mathbf{x}, \tau) &= b_\delta \delta_{n\text{LPT}}(\mathbf{x}, \tau) \\ &+ b_{\sigma^2} [\sigma_\Lambda^{(1)}]^2(\mathbf{x}, \tau) + b_{\text{tr}[M^{(1)}M^{(1)}]} \text{tr} \left[M_\Lambda^{(1)} M_\Lambda^{(1)} \right](\mathbf{x}, \tau) \\ &+ b_{\nabla^2 \delta} \nabla^2 \delta_{n\text{LPT}}(\mathbf{x}, \tau), \end{aligned} \quad (3.11)$$

where we emphasize the presence of the cutoff Λ . The linear displacement tensor $M_\Lambda^{(1)}$ is related to the linear density perturbation via

$$M_{ij,\Lambda}^{(1)}(\mathbf{q}) = -\frac{\partial_{q_i} \partial_{q_j}}{\nabla_q^2} \delta_\Lambda^{(1)}(\mathbf{q}),$$

where $\delta_\Lambda^{(1)}$ is defined in Eq. (3.4).

3.2.3 Field-level likelihood

Instead of modeling directly the tracer number counts, as done for example in [20, 23], the EFT likelihood aims to describe the tracer density field δ_d . It is obtained by integrating out the modes with $k > \Lambda$ in the initial conditions [110]. These small-scale modes also produce a stochastic contribution to the predicted galaxy density field $\delta_{\text{det},\Lambda}$. This effect is encoded by the noise field ϵ . Since this effective noise field arises from the superposition of many independent modes, it is Gaussian to leading order. Moreover, because the tracer formation is spatially local, the power spectrum of the noise is constant to leading order, with corrections scaling as k^2 as shown already in Section 2.4 (see also [31]). In other words,

$$\langle \epsilon(\mathbf{k}) \epsilon(\mathbf{k}') \rangle = (2\pi)^3 \delta_D(\mathbf{k} + \mathbf{k}') P_\epsilon (1 + \sigma_{\epsilon,2} k^2 + \dots), \quad (3.12)$$

where δ_D denotes the Dirac delta function, and we have denoted with $P_\epsilon \sim \bar{n}_d^{-1}$ the leading order contribution to the noise covariance, i.e. the leading noise power spectrum. Note that the scale-dependent correction $\sigma_{\epsilon,2}$ is written here as fractional correction by convention. In general, the gravitational evolution of small-scale modes under the influence of large-scale modes generates density-dependent noise terms, which cause the noise covariance to be non-diagonal in Fourier space. We do not consider these contributions here, since all synthetic datasets used here – except for 2LPT with second-order bias and

cutoff mismatch (between synthetic data and inference) – do not contain such density-dependent noise. We also set the subleading noise contribution $\sigma_{\epsilon,2}$ to zero throughout this chapter. More detailed investigations of the impact of the noise model are relegated to future work.

Thus, our final data model is given by

$$\begin{aligned}\delta_{d,\Lambda}(\mathbf{x}) &= \delta_{\text{det},\Lambda}(\mathbf{x}) + \epsilon(\mathbf{x}), \\ \epsilon(\mathbf{x}) &\leftarrow \mathcal{G}(\epsilon; 0, \sigma_\epsilon^2 \mathbb{1}),\end{aligned}\tag{3.13}$$

where $\delta_{\text{det},\Lambda}$ is given by one of the forward models we consider here, namely Eqs. (3.3)–(3.5) or Eq. (3.11). In the second line of 3.13 we emphasize that the noise field is assumed to be Gaussian distributed with covariance structure given by the leading term from Eq. (3.12). Specifically, the noise variance σ_ϵ^2 on the real-space grid of size N_g^Λ is defined through the following relation

$$P_\epsilon = \sigma_\epsilon^2 \frac{L^3}{(N_g^\Lambda)^3}.\tag{3.14}$$

We consider two forms of EFT likelihoods: one whose arguments explicitly include bias parameters, hence labeled *unmarginalized likelihood*, and the other one where bias parameters are analytically marginalized out, namely the *marginalized likelihood*. Below we detail the two likelihoods in that order. The notation closely follows that in [104]. Further, we switch to the discrete Fourier space representation of the fields (see Appendix 3.7 for details on our Fourier space convention).

The expression for the unmarginalized likelihood is given by a Gaussian following our assumptions on the noise field discussed at the beginning of this section. Explicitly,

$$\begin{aligned}\ln \mathcal{L} \left(\delta_{d,\Lambda} \middle| \alpha, \hat{s}, \{b_O\}, \sigma_\epsilon \right) &= -\frac{1}{2} \sum_{\mathbf{k} \neq 0}^{k_{\max}} \left[\ln 2\pi\sigma_\epsilon^2 + \frac{1}{\sigma_\epsilon^2} \left| \delta_{d,\Lambda}(\mathbf{k}) - \delta_{\text{det},\Lambda}[\alpha, \hat{s}, \{b_O\}](\mathbf{k}) \right|^2 \right] \\ &+ \text{const.},\end{aligned}\tag{3.15}$$

where we have explicitly stated the dependence of $\delta_{\text{det},\Lambda}$ on bias parameters, $\{b_O\}$, the scaling parameter α and the initial conditions \hat{s} . Note that $\delta_{\text{det},\Lambda}$ is a deterministic function of these parameters, and that the likelihood is evaluated only up to k_{\max} , strictly allowing only for modes below the cutoff. In order to maximize the information gain, we choose $k_{\max} = \Lambda$. We accumulate all the terms which depend neither on α , $\{b_O\}$, \hat{s} nor σ_ϵ in const., since they represent only irrelevant normalizing factors.

The marginalized likelihood is obtained by marginalizing analytically over all bias parameters in Eq. (3.15). Given that the likelihood depends quadratically on any bias coefficient (in case priors on the bias parameters are gaussian or uniform on $(-\infty, \infty)$),

this is a straightforward calculation (see section 2.2 in [104]). The expression is

$$\begin{aligned}
-\ln \mathcal{L} \left(\delta_{d,\Lambda} \middle| \alpha, \hat{s}, \sigma_\epsilon \right) &= \frac{1}{2} \text{tr} \ln A_{OO'} + \frac{1}{2} \text{tr} \ln C_{\text{prior}} + \frac{1}{2} \sum_{\mathbf{k} \neq 0}^{k_{\text{max}}} \left(\ln \sigma_\epsilon^2 + \frac{1}{\sigma_\epsilon^2} |\delta_{d,\Lambda}(\mathbf{k})|^2 \right) \\
&\quad - \frac{1}{2} \sum_{\{O,O'\}} B_O (A^{-1})_{OO'} B_{O'} + \text{const.} , \\
B_O \equiv B_O[\hat{s}, \alpha] &= \sum_{\mathbf{k} \neq 0}^{k_{\text{max}}} \frac{\delta_{d,\Lambda}(\mathbf{k}) O[\hat{s}, \alpha]^*(\mathbf{k})}{\sigma_\epsilon^2} + \sum_{O'} (C_{\text{prior}}^{-1})_{OO'} \mu_{b_{O'}} , \\
A_{OO'} \equiv A_{OO'}[\hat{s}, \alpha] &= \sum_{\mathbf{k} \neq 0}^{k_{\text{max}}} \frac{O[\hat{s}, \alpha](\mathbf{k}) O'[\hat{s}, \alpha]^*(\mathbf{k}')}{\sigma_\epsilon^2} + (C_{\text{prior}}^{-1})_{OO'} , \tag{3.16}
\end{aligned}$$

where, once again, const. encapsulates the terms independent of the parameters of interest. The marginalization was performed under the assumption of a Gaussian prior on bias parameters with covariance C_{prior} and mean μ_{b_O} . We choose a fairly uninformative prior as given in Eq. (3.19). The information on α and the initial conditions \hat{s} is propagated through the B_O and $A_{OO'}$ operators.

3.3 Sampling the full posterior

Here, we provide the final expression for the posterior being sampled and elaborate more on some specific choices of our sampling scheme, as well as some additional details of our code implementation.

3.3.1 Full posterior

The results of the previous section now allow us to write the full posterior which we aim to sample from. The expression can be obtained readily from

$$\mathcal{P} \left(\alpha, \hat{s}, \{b_O\}, \sigma_\epsilon \middle| \delta_{d,\Lambda} \right) = \frac{\mathcal{L}(\delta_{d,\Lambda} | \alpha, \hat{s}, \{b_O\}, \sigma_\epsilon) \mathcal{P}(\alpha, \hat{s}, \{b_O\}, \sigma_\epsilon)}{\mathcal{P}(\delta_{d,\Lambda})} , \tag{3.17}$$

where $\mathcal{P}(\alpha, \hat{s}, \{b_O\}, \sigma_\epsilon | \delta_{d,\Lambda})$ represents the posterior probability of the parameters of interest given the synthetic data, $\mathcal{L}(\delta_{d,\Lambda} | \alpha, \hat{s}, \{b_O\}, \sigma_\epsilon)$ the corresponding likelihood (see Eqs. (3.15)–(3.16)), while $\mathcal{P}(\alpha, \hat{s}, \{b_O\}, \sigma_\epsilon)$ represents the associated prior. The $\mathcal{P}(\delta_{d,\Lambda})$ represents the evidence, which doesn't play any role in our inference framework. As noted in Section 3.1, we keep all other cosmological parameters fixed to the fiducial values listed there.

We assume minimal prior knowledge on σ_ϵ and the $\{b_O\}$. Moreover, as physical parameters describing the properties of the tracers, they are a-priori independent of the

initial conditions \hat{s} . Therefore, the joint prior structure is entirely factorized. We choose the following prior configuration throughout for the unmarginalized likelihood:

$$\begin{aligned}\mathcal{P}(\sigma_\epsilon) &= \mathcal{U}(0.05, 100.), \\ \mathcal{P}(\alpha) &= \mathcal{U}(0.5, 1.5), \\ \mathcal{P}(b_\delta) &= \mathcal{U}(0.01, 10), \quad \mathcal{P}(b_{\delta^2}) = \mathcal{U}(-10, 10), \\ \mathcal{P}(b_{\sigma^2}) &= \mathcal{P}(b_{\text{tr}[M^{(1)}M^{(1)}]}) = \mathcal{P}(b_{\nabla^2\delta}) = \mathcal{U}(-25, 25),\end{aligned}\tag{3.18}$$

where $\mathcal{U}(l, r)$ denotes the uniform distribution on interval $[l, r]$. Note that here we choose to keep the prior on b_δ strictly positive, as we do not consider the modeling of negative bias tracers here, such as voids. We chose priors of the higher-order bias coefficients to be symmetric around zero, as a-priori these bias coefficients could take on either sign.

For the inference with the marginalized likelihood, the only difference is in the priors for the bias parameters, which are taken to be of the following form

$$\begin{aligned}\mathcal{P}(b_\delta) &= \mathcal{G}(0.01, 10); \quad \mathcal{P}(b_{\delta^2}) = \mathcal{G}(0, 10) \\ \mathcal{P}(b_{\sigma^2}) &= \mathcal{P}(b_{\text{tr}[M^{(1)}M^{(1)}]}) = \mathcal{P}(b_{\nabla^2\delta}) = \mathcal{G}(0, 25).\end{aligned}\tag{3.19}$$

Given the signal-to-noise level of our synthetic datasets, these priors are essentially uninformative, and we expect a entirely negligible difference in parameter inferences between marginalized and unmarginalized likelihoods.

Finally, for the prior $\mathcal{P}(\hat{s})$ on the initial conditions, we consider the following two choices:

$$\mathcal{P}(\hat{s}) = \begin{cases} \delta_D(\hat{s} - \hat{s}_{\text{true}}) & \text{for FIXEDIC case} \\ \mathcal{G}(\hat{s}; 0, S) & \text{for FREEIC case,} \end{cases}\tag{3.20}$$

where we have separated the cases with fixed initial conditions to the ground-truth (FIXEDIC) and with initial conditions explicitly sampled (FREEIC). In the latter case, the prior covariance structure of the discretized \hat{s} field is given by $S^{\mathbf{x}_i \mathbf{x}_j} \equiv \langle \hat{s}(\mathbf{x}_i) \hat{s}(\mathbf{x}_j) \rangle_{\mathcal{P}(\hat{s})} = \delta_D^{i,j}$, while in Fourier space it becomes $S^{\mathbf{k} \mathbf{k}'} \equiv \langle \hat{s}(\mathbf{k}) \hat{s}(\mathbf{k}') \rangle_{\mathcal{P}(\hat{s})} = (N_g^\Lambda)^3 \delta_D^{\mathbf{k}, \mathbf{k}'}$, with N_g^Λ being the grid size corresponding to the cutoff Λ , and δ_D denoting the Kronecker delta (see Appendix 3.7 for details on our Fourier convention).

The FIXEDIC case corresponds to that considered in the application to dark matter halo catalogs in previous papers of this series [30, 102, 104, 106, 107], where unbiased inference of α was shown. Moreover, the same setup was used to measure bias parameters of simulated halos and galaxies in [118, 119]. For real-data applications, as we have no knowledge of the true initial conditions a-priori, it is crucial for our method to properly marginalize over all plausible realizations of the initial conditions. Thus, the parameter posterior obtained on our synthetic datasets with the FIXEDIC prior serves as a good reference point for the FREEIC case. Specifically, we expect consistency between the FIXEDIC and FREEIC posterior means of the parameters, which we demonstrate in Section 3.5.

3.3.2 Sampling methods

In order to explore the posterior surface of the parameters of our model, we utilize a combination of slice sampling (see Section 29.7 in [86] and [91]) and Hamiltonian Monte Carlo (HMC) sampling techniques (see [88, 89] and Section 30.1 in [86]). Below, we motivate this choice of sampling techniques, which was inspired by the findings made in the development of the BORG code (see, e.g. [23]). See also Section 3.3.2 for complementary text on HMC and slice sampling.

The slice sampling technique² is used for sampling α , bias parameters $\{b_O\}$, and the noise parameter σ_ϵ . This means that we actually sample the 1D probability density function of these parameters conditioned on the current realization of initial conditions \hat{s} . We adopt sequential univariate slice sampling, i.e. we sequentially sample individual parameters. While multivariate slice sampling methods do exist, they require additional tuning to be more efficient than our approach (see the discussion in section 5 of [91]).

When it comes to sampling the posterior of the initial conditions \hat{s} , we use the HMC sampling technique. The reason behind this choice is that the number of Monte Carlo samples needed for generating an independent sample scales more efficiently with the problem dimensionality than for standard Monte Carlo methods. This scaling goes as $\sim N_{\text{dim}}$ in the case of standard random-walk algorithm, while it goes as $\sim N_{\text{dim}}^{1/4}$ for the HMC method (see, for example, Section 4.4 of [88]). For the problem we consider here, typically $N_{\text{dim}} \sim 10^5 - 10^6$. Therefore, HMC currently appears to be the most (if not only) practical sampling method to tackle such a problem.

In order to utilize HMC, one needs a fully differentiable forward model with respect to the initial conditions. This requirement is necessary for the crucial step of generating a new proposal of initial conditions, consistent with the data likelihood. This new proposal is generated by numerically integrating along the Hamiltonian flow defined by the likelihood and prior gradients with respect to the initial conditions \hat{s} . For this, we choose the second-order leapfrog integrator, although see, e.g. [120, 121], for the applicability of higher-order integration schemes.

Given the structure of our code, `LEFTfield`³, described in Section 3.3.3, the analytical derivative of the full forward model can be readily obtained through successive applications of the chain rule.

Finally, readers may ask why not try to combine the two sampling approaches such that all the parameters are sampled within the HMC scheme at the same time. This however is unfeasible, due to the large difference between the derivative norms of the variation with respect to $\hat{s}(\mathbf{k})$ and the variation of the rest of the model parameters. Variation of α or bias parameters affects all the modes of the forward model, while the derivative with respect to a given $\hat{s}(\mathbf{k})$ only has to vary a single mode at a time. This in turn requires the HMC trajectories to be integrated with small step sizes, in order to

²See also Algorithm 1

³*Lagrangian, Effective-Field-Theory-based forward model of cosmological density fields.*

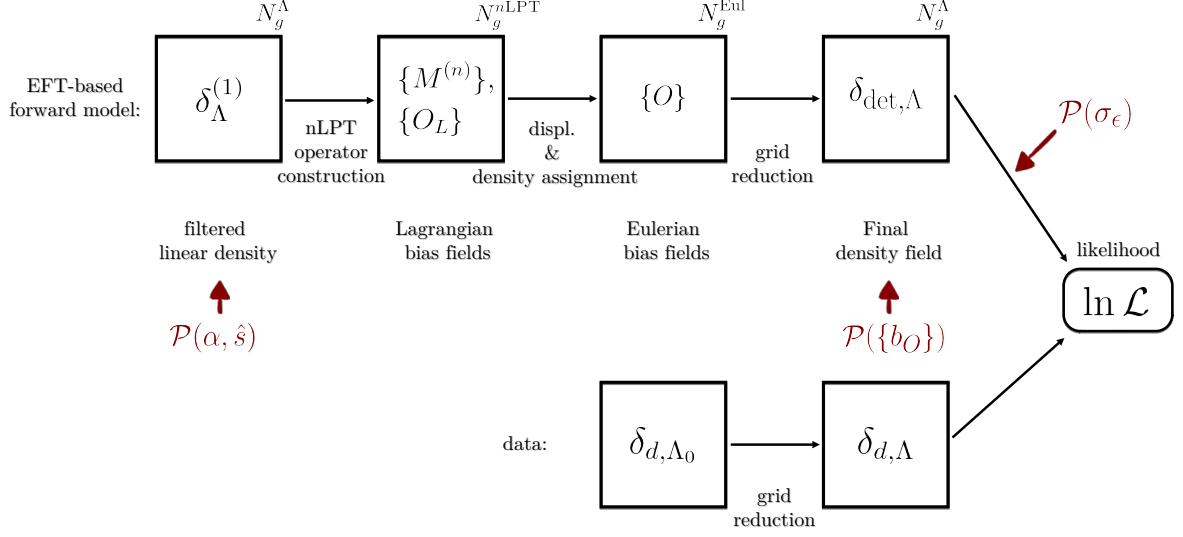


Figure 3.1: A depiction of our consistency test architecture, which consists of two branches. The top one describes all the relevant forward elements of the forward models we consider. All forward model elements are described in Section 3.2 and Section 3.3.3. The bottom row schematically shows the synthetic data generation process. Note that we use the same forward models from the top row to generate the dataset in the bottom row, but at a different cutoff value Λ_0 . Our ensemble of synthetic datasets is described in Section 3.4.

reach a reasonable acceptance rate, and results in a very slow exploration of the (joint) posterior surface.

Therefore, decoupling the sampling of α , bias and noise parameters from the sampling of initial conditions seems to provide the fastest exploration time. Note that, in such a block-sampling scheme one could still use the HMC method to sample cosmological and bias parameters, conditioned on a given realization of the initial conditions. That is, one could separate the former and latter into two separate HMC sampling blocks with two different equations of motion (to be integrated). However, since the dimension of the cosmological and bias parameter space is currently negligible, we opt for the robustness of slice sampling.

3.3.3 Code implementation

The forward models and likelihoods described in Section 3.2 are implemented in the differentiable code `LEFTfield`. `LEFTfield` adopts the C++17 standard and represents a substantially extended and more efficient version of the code presented in [30]. Most importantly, `LEFTfield` implements the gradient of the likelihood with respect to the initial conditions \hat{s} , which as noted is crucial for the HMC sampling approach. A public release of the code, subject to additional tests and tidying up, is planned for the future.

The code is structured in a modular way, breaking down the forward model into a series of simple steps, called “forward elements”, which are templated on generic input and output types and implement the general behavior of composed operator chains; specifically, the respective input and output types need to match for all composed forward elements. The sequence of high-level operations contained in the n LPT forward models are represented in the top row of Fig. 3.1 (note that each of the blocks in general consists of multiple forward elements). The forward model starts from a sample of initial conditions from which one obtains the initial density field $\delta_{\Lambda}^{(1)}$, shown in top left of Fig. 3.1. The size of this grid is chosen based on the cutoff Λ . This is indicated with N_g^{Λ} . The next element is the bias operator construction. For this, we use the set of bias operators appearing in Eq. (3.10) in case of 1LPT and 2LPT models. At this step, care is also taken for representing all the physical modes of the forward model, by choosing appropriate grid sizes, indicated with $N_g^{n\text{LPT}}$. Afterwards, these fields are displaced utilizing a weighted particle scheme to the final Eulerian positions. This results in a mapping $O_L \rightarrow O$, where now the O operators are assigned onto a grid of size N_g^{Eul} , chosen in advance, with $N_g^{\text{Eul}} \geq N_g^{n\text{LPT}}$ in order to keep all physical modes of the forward model represented on the Eulerian grid. At the end, the set of the displaced bias operators $\{O\}$ is resized in Fourier space to a smaller and final grid corresponding to N_g^{Λ} , using the sharp- k cutoff. Finally, the deterministic prediction $\delta_{\text{det},\Lambda}$ from Eq. (3.11) is constructed in the last piece of the top row. This also involves drawing a new set of relevant bias parameters from the corresponding prior. The grid reduction and grid-padding are both performed in Fourier space. Several options for mass assignment schemes are implemented, including nearest-grid-point (NGP, which is not differentiable however), cloud-in-cell (CIC), and triangular-shaped cloud (TSC). Here, we use the CIC scheme throughout. The very last piece of the forward model is the evaluation of the likelihood, given by either Eq. (3.15) or Eq. (3.16). Note that the forward models of Section 3.2.1 are much simpler, but have the same overall structure.

In case of the HMC sampling block, the full gradient of the likelihood with respect to the initial conditions, \hat{s} , needs to be evaluated. In order to do so, we utilize the chain rule, collecting every \hat{s} -dependent term of each element in the forward model, from right to left in the flowchart. In addition, for every sample of \hat{s} , the slice sampler generates a new sample of the other parameters of interest. This process is repeated until the desired number of samples is achieved (see Section 3.5 and Appendix 3.10 for more details on

Dataset \ Parameter	α	b_δ	b_{δ^2}	b_{σ^2}	$b_{\text{tr}[M^{(1)}M^{(1)}]}$	$b_{\nabla^2\delta}$	σ_ϵ	$\tilde{\sigma}_\epsilon$
$\mathcal{D}_1^{\text{LINEAR}}$	1.0	1.0	0	0	0	0	0.5	0.586
$\mathcal{D}_2^{\text{LINEAR}}$	1.0	1.0	1.0	0	0	0	0.5	0.586
$\mathcal{D}_2^{\text{LINEAR-informative}}$	1.0	1.0	1.0	0	0	0	0.001	0.002
$\mathcal{D}_1^{2\text{LPT}}$	1.0	1.0	0	0	0	0	0.5	0.586
$\mathcal{D}_2^{2\text{LPT}}$	1.0	0.87	0	-0.2	-0.2	0.2	0.4	0.469

Table 3.1: Parameters used for generating different synthetic datasets. The noise levels in these datasets correspond to $P_\epsilon = 2743 h^{-3} \text{Mpc}^3$, except for the dataset from third row, which has $P_\epsilon = 1.1 \times 10^{-2} h^{-3} \text{Mpc}^3$. All datasets listed in this table are obtained using a cutoff of $\Lambda_0 = 0.14 h \text{Mpc}^{-1}$. The parameters listed in this table are described in Section 3.2, except $\tilde{\sigma}_\epsilon$, which is defined in Eq. (3.21).

our convergence requirements and verification).

The bulk of the computing time is spent in the HMC sampling of the initial conditions. For reference, we provide some benchmark computing times here. For the 2LPT forward model with 64^3 grid size, `LEFTfield` generates ~ 200 samples per CPU hour, roughly corresponding to ~ 1 effective sample per CPU hour, running on a single Intel(R) Xeon(R) Gold 6138 CPU @ 2.00GHz with 20 cores and using `OpenMP` parallelization.

3.4 Synthetic datasets

In this section, we describe how precisely we generate the synthetic data sets (the first element in the bottom row of Fig. 3.1). In general, a dataset is generated from each of the aforementioned forward models. We introduce model mis-specification through the mismatch between the cutoff Λ_0 in the synthetic data and a varying cutoff Λ in our forward models, as indicated in the two elements in the bottom row of Fig. 3.1. We always fix $\Lambda_0 = 0.14 h \text{Mpc}^{-1}$. Throughout, we label the specific realization of initial conditions used for synthetic data generation by \hat{s}_{true} . All parameters of the synthetic datasets are summarized in Tab. 3.1.

3.4.1 LINEAR model synthetic data

For the models described in Eqs. (3.3)–(3.5) we generate two sets of synthetic data: one with $b_{\delta^2} = 0$ and the other with $b_{\delta^2} \neq 0$. Below, we specify the parameter values adopted and explain our choice.

The first case of synthetic data, $\mathcal{D}_1^{\text{LINEAR}}$, is obtained with the parameters listed in the first row of Tab. 3.1. Note that the cutoff set by Λ_0 determines the grid size, which in this case is $N_g^{\Lambda_0} = 90$, given that it corresponds to the Nyquist frequency for the cutoff $\Lambda_0 = 0.14 h \text{Mpc}^{-1}$ and a box size $L = 2000 h^{-1} \text{Mpc}$. The parameter $\sigma_\epsilon = 0.5$ is the

square root of the noise variance on this grid, which corresponds to a Poisson shot-noise for tracers with comoving number density of $\bar{n} \approx 3.645 \times 10^{-4} (h \text{ Mpc}^{-1})^3$. Since the noise power spectrum P_ϵ is a physical quantity (in particular, independent of the grid size), it follows from Eq. (3.14) that the combination $\sigma_\epsilon (N_g^\Lambda)^{-3/2}$ must be independent of the grid size corresponding to the cutoff Λ . This implies that σ_ϵ itself will show a cutoff-dependence induced purely by the changing grid size N_g^Λ . Therefore, instead of working with σ_ϵ , we define the following quantity

$$\tilde{\sigma}_\epsilon \equiv 10^3 \frac{\sigma_\epsilon}{(N_g^\Lambda)^{3/2}} = 10^3 \left(\frac{P_\epsilon}{L^3} \right)^{1/2}, \quad (3.21)$$

which is grid-size-independent by construction. The prefactor 10^3 is introduced for numerical convenience. We will mainly quote $\tilde{\sigma}_\epsilon$ instead of σ_ϵ in our posterior analyses in Section 3.5. As can be seen from Tab. 3.1, we adopt a comparable noise level for all synthetic datasets except for $\mathcal{D}_2^{\text{LINEAR-informative}}$, which we describe below.

The second synthetic dataset, $\mathcal{D}_2^{\text{LINEAR}}$ comes in two variants, listed in the second and third row of Tab. 3.1. We always choose $|b_{\delta^2}| = |b_\delta|$ in order to introduce a non-negligible mode coupling through the quadratic term in Eq. (3.5). In addition, we consider $\mathcal{D}_2^{\text{LINEAR-informative}}$ (third row of Tab. 3.1), using a very low noise level ($\tilde{\sigma}_\epsilon = 0.002$) and hence representing a very informative dataset. This dataset was included to further investigate the dependence of the inferred b_{δ^2} as a function of the cutoff Λ (see Section 3.5.1).

Apart from the case of $\mathcal{D}_1^{\text{LINEAR}}$, for all other datasets, including $\mathcal{D}_2^{\text{LINEAR}}$ datasets, we generate two different data realizations. We achieve this by generating two different initial conditions realizations, keeping the values for the remaining parameters fixed. Independent inferences are performed on both data realizations. This helps us gauge the significance of any mis-estimation of the posterior and hence of potential systematic trends in the inferred parameters. Henceforth, we label the different data realizations by the subscript $\{a, b\}$, for example $\mathcal{D}_{2,a}^{\text{LINEAR}}$ and $\mathcal{D}_{2,b}^{\text{LINEAR}}$ which correspond to the two different realizations of the synthetic dataset listed in the second row of Tab. 3.1.

3.4.2 2LPT synthetic data

We generate two types of synthetic datasets for the 2LPT forward model described in Eq. (3.11). They are labelled as $\mathcal{D}_1^{2\text{LPT}}$ and $\mathcal{D}_2^{2\text{LPT}}$ and their parameters are listed in the last two rows of Tab. 3.1. As before, we also generate two different realizations of each of these datasets.

The $\mathcal{D}_1^{2\text{LPT}}$ datasets serve as an input for the internal consistency between the 2LPT and 1LPT forward models. In particular, these correspond to a noisy, but unbiased tracer of the matter field itself, given that $b_\delta = 1$ with all higher-order bias coefficients set to zero. We demonstrate in Section 3.5.2 that we exactly recover the fiducial values

of parameters in case of 2LPT and that we also recover the expected shifts of parameters in the case of the 1LPT forward model. We calculate these shifts analytically in Appendix 3.11.

Finally, the $\mathcal{D}_2^{2\text{LPT}}$ datasets contain nonzero higher-order bias coefficients. These are the most realistic datasets considered here, in the sense that they contain the complications due to both nonlinear gravity and nonlinear bias. Since the calculation of the running of the bias parameters with cutoff Λ is substantially more involved in this case, we instead validate our field-level forward model based on the inference of the α parameter for this case. That is, if the EFT likelihood is able to correctly absorb the effect of modes above the cutoff, then it should lead to an unbiased inference of α for all values of Λ .

3.5 Consistency test results

In this section, we describe the analysis procedure of our MCMC chains. Since the dimensionality of our posterior is exceptionally high, fully characterizing it is challenging [122], even with HMC sampling. To ensure that our samples fairly represent the true posteriors, we strictly adopt the following setup and procedure:

- We run MCMC chains with both free initial conditions (FREEIC) and initial conditions fixed to the ground-truth (FIXEDIC). Since it is expected that the posterior of the FIXEDIC case is within the typical set of the FREEIC posterior, in case of no strong multi-modality, it serves as a good reference point. Indeed, we find a good agreement between the joint posteriors of these two cases (see also the next bullet point), suggesting no strong multi-modality is present in the tests we consider in this work.
- For FREEIC runs, we run at least three chains: two starting from randomized values of initial conditions and sampled parameters, and one more chain starting from the ground-truth. The latter serves as an additional check on multimodality of the posterior, and of the convergence of our chains. The remaining parameters differ among the forward models we consider here and we always indicate which parameters are actually sampled.
- In our analysis, we discard the initial part of each chain, which is typically 5 – 10 correlation lengths long (see Appendix 3.10 on how we obtain the correlation lengths). Throughout, for each reported inference, if we run more than one MCMC chain as described in the previous point, we combine the chains into one single set of posterior contours. The consistency between different chains being combined is verified with the Gelman-Rubin statistics described in the next point.

- We evaluate the Gelman-Rubin statistics [123–125] for our MCMC chains as described in Appendix 3.10. In doing so, we also quantify the (combined) effective sample size. The results for both, for all our chains, are listed in Tab. 3.2, 3.3 and 3.4. We require all of our chains to have $\gtrsim 100$ effective samples. This allows us to have the MCMC sampling error reduced to $\sim 10\%$, which is sufficient for the purposes of this work.

It is also important to note that for the FREEIC chains, we use different k -binned quantities in order to check their statistics and convergence. These are the power spectrum of \hat{s} , the mean deviation from \hat{s}_{true} and the corresponding power spectrum of this deviation. Additionally, we have verified that the convergence of individual \hat{s} modes is well represented by that of the k -binned quantities for the different bins. We also note that both the k -binned \hat{s} quantities and the individual \hat{s} modes converge much faster than other parameters of the model, namely α , $\{b_O\}$ and σ_ϵ .

We will also compare the sampled \hat{s} posterior with analytical predictions. For the latter, we always first calculate the per- k -mode prediction, and then compute the k -bin average, which is then compared with the corresponding sampled posterior in the same k -bin.

3.5.1 LINEAR forward models

In this section, we focus on the forward models described by Eq. (3.3) and Eq. (3.5).

Linear bias

First, we discuss the results of the forward model with linear bias expansion. In this case, it is possible to calculate the posterior of the initial conditions analytically. The comparison with the sampled posterior then verifies whether our sampling approach indeed fully explores the posterior in this case. As a first result, we focus on Fig. 3.2. In this figure, we show the projection of the posterior to the $b_\delta - \tilde{\sigma}_\epsilon$ plane. We distinguish two cases. The case where the posterior contours were obtained by fully marginalizing over the initial conditions (FREEIC), shown in the left panel, and the case where the initial conditions were fixed to the ground-truth (FIXEDIC), shown in the right panel. The two panels indicate that the FREEIC and FIXEDIC posterior means are consistent. Note the stark contrast in the posterior widths between the FIXEDIC and FREEIC cases. This is explained by the fact that, in the case of FIXEDIC, only two parameters need to be constrained, while in case of FREEIC the joint posterior simultaneously constrains $\gtrsim 10^5$ degrees of freedom. More specifically, the free initial conditions also allow for an overall change in the amplitude, leading to the wider posteriors in b_δ . We also observe that the posterior contours shrink with increasing Λ , as expected. The degeneracy between the amplitudes of the signal ($\propto b_\delta$) and noise ($\propto \tilde{\sigma}_\epsilon$) is harder to break at lower Λ due to the

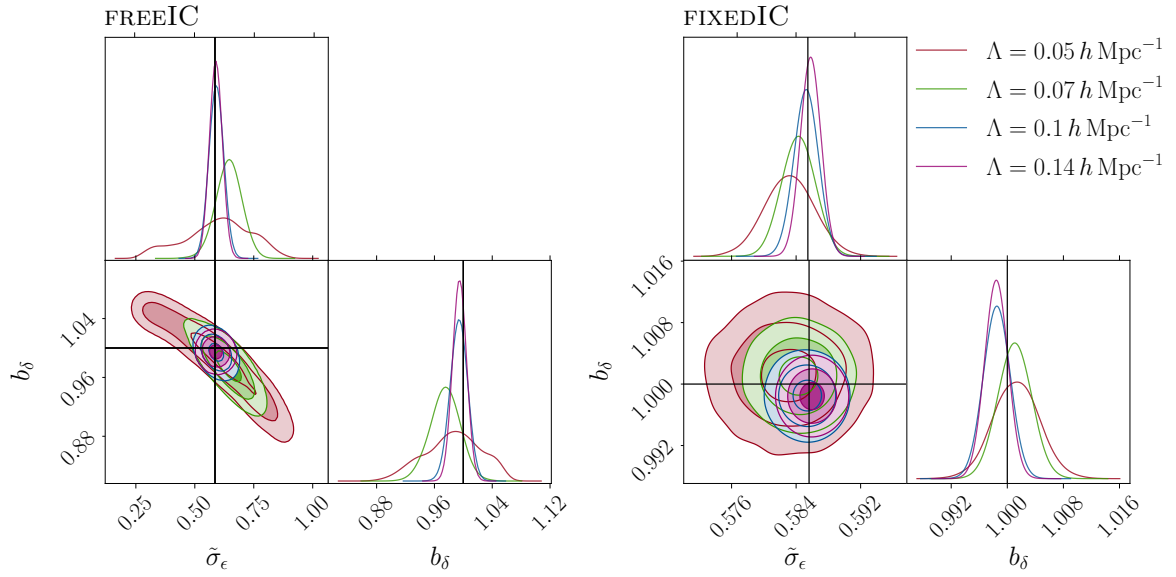


Figure 3.2: Inferred posterior of the parameters of the forward model from Eq. (3.3) on the $\mathcal{D}_1^{\text{LINEAR}}$ synthetic dataset. On the left, the posterior projections to the $\tilde{\sigma}_\epsilon - b_\delta$ plane for the FREEIC case and different cutoffs are shown, while on the right we show the corresponding case of FIXEDIC posteriors. Note the difference in the axis ranges between the FIXEDIC and FREEIC posteriors. The black lines indicate expected values for both parameters. As can be seen, the FREEIC and FIXEDIC posteriors agree with each other. Also, note the difference in the posterior uncertainties between forward models with different cutoff.

shallower slope of the linear power spectrum, hence resulting in posterior uncertainties that grow faster toward smaller Λ than expected merely from mode counting arguments; that is, the error bar on b_δ grows toward smaller Λ more rapidly than $\Lambda^{-3/2}$. The stronger degeneracy in the $b_\delta - \tilde{\sigma}_\epsilon$ plane also results in a slower exploration by the samplers, evidenced by a longer correlation length, which we do not explicitly show here for conciseness.

Another point to emphasize is that the correct fiducial noise level has also been recovered at 68 – 95%CL in all cases. This means that our inferences have been able to clearly disentangle between the actual signal and Gaussian noise contributions.

We now turn to investigating the posterior of the initial conditions \hat{s} , especially to see if the inference recovers the true initial conditions \hat{s}_{true} . Fig. 3.3 compares the inferred and true initial conditions as a function of Fourier wave number k , in bins of k . The inference was performed on the $\mathcal{D}_1^{\text{LINEAR}}$ dataset, at the cutoff of $\Lambda = 0.1 h \text{ Mpc}^{-1}$, keeping the other model parameters $\{b_\delta, \tilde{\sigma}_\epsilon\}$ free. The top panel of Fig. 3.3 shows the ratio between the power spectrum of \hat{s} and that of \hat{s}_{true} . The ratio is consistent with unity, indicating that the two fields agree in terms of power, or mean amplitude.

The middle panel of Fig. 3.3 depicts the k -bin statistics of the residuals $\Delta_{\hat{s}}(\mathbf{k}) = (\hat{s} - \hat{s}_{\text{true}})(\mathbf{k})$. Indeed, their distribution is centered around 0, clearly implying that the bulk of the \hat{s} posterior closely traces the \hat{s}_{true} field.

Note that we do not expect the posterior to be always centered around the \hat{s}_{true} field, but that \hat{s}_{true} is within the typical set. We can in fact be more precise. As we show in Appendix 3.8.1, the \hat{s} posterior mean and covariance for the linear model considered here is, *in the case when the parameters b_δ and $\tilde{\sigma}_\epsilon$ are fixed to their fiducial values*, given by

$$\begin{aligned}\hat{s}_{\text{WF}} &= C_{\text{WF}} R^T C_\epsilon^{-1} \delta_{d,\Lambda}, \\ C_{\text{WF}} &= (1 + S R^\dagger C_\epsilon^{-1} R)^{-1} S,\end{aligned}\tag{3.22}$$

with

$$\begin{aligned}R_{\mathbf{k}_2}^{\mathbf{k}_1} &= \delta_D^{\mathbf{k}_1, \mathbf{k}_2} b_\delta T(k_1) \\ (C_\epsilon)_{\mathbf{k}_2}^{\mathbf{k}_1} &= \delta_D^{\mathbf{k}_1, \mathbf{k}_2} P_\epsilon \\ S_{\mathbf{k}_2}^{\mathbf{k}_1} &= \delta_D^{\mathbf{k}_1, \mathbf{k}_2} (N_g^\Lambda)^3,\end{aligned}\tag{3.23}$$

and $\delta_{d,\Lambda}$ the reduced density field of the $\mathcal{D}_1^{\text{LINEAR}}$ synthetic dataset. As seen from Eq. (3.22), the \hat{s} posterior exhibits two limits. First, in the limit of uninformative data, i.e. large noise C_ϵ , the posterior simply approaches the prior, and the posterior mean of \hat{s} approaches zero while $C_{\text{WF}} \rightarrow S$. Second, in the limit of very informative data, i.e. small noise C_ϵ , the posterior mean and covariance approach $\hat{s} \rightarrow R^{-1} \delta_{d,\Lambda} \rightarrow \hat{s}_{\text{true}}$, following Eq. (3.3), and $C_{\text{WF}} \rightarrow (S R^\dagger C_\epsilon^{-1} R)^{-1} S \rightarrow 0$, respectively.

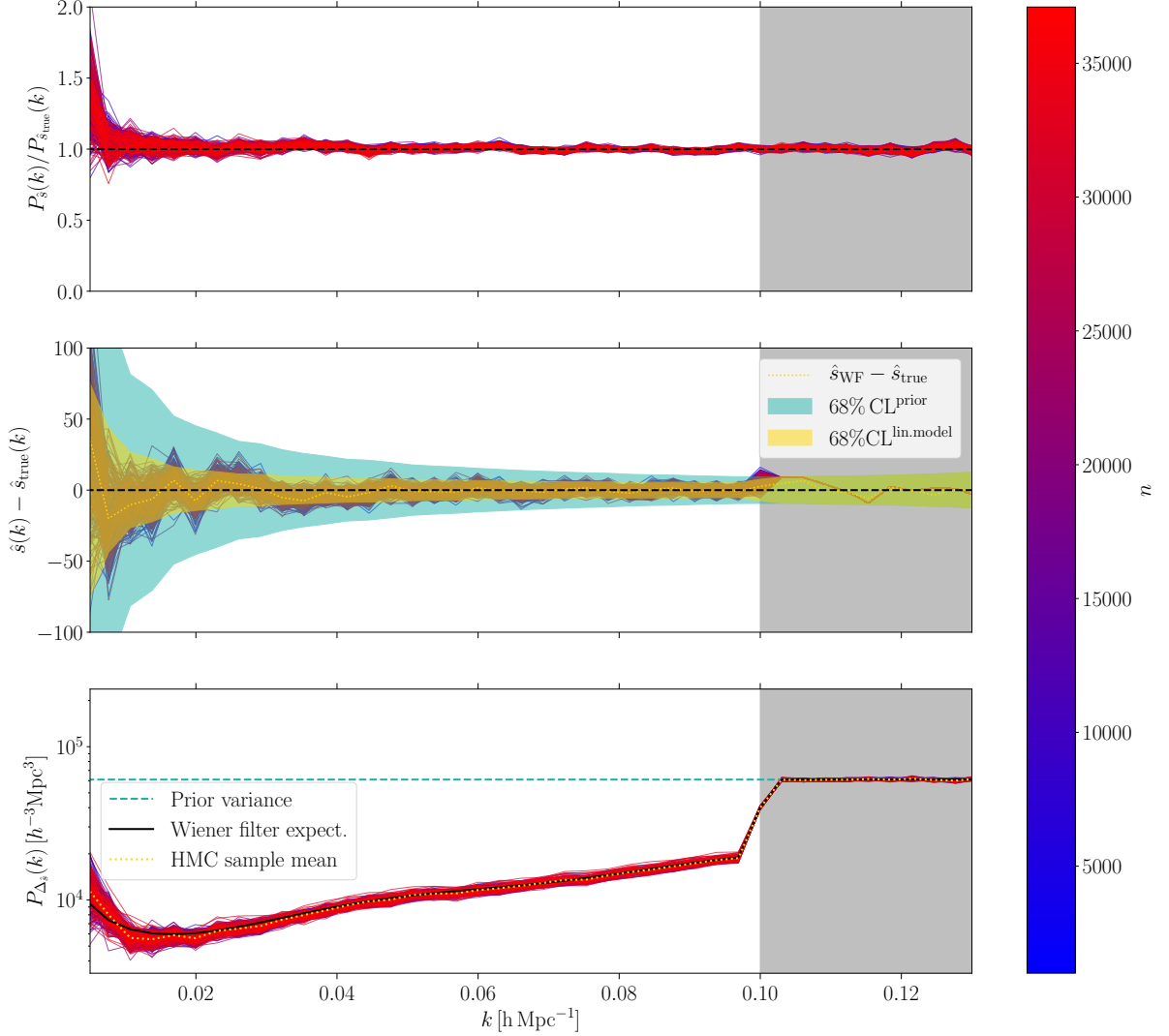


Figure 3.3: Samples of different binned statistics of the inferred \hat{s} field, using $\Lambda = 0.1 h \text{ Mpc}^{-1}$, for the linear forward model from Eq. (3.3) on the $\mathcal{D}_1^{\text{LINEAR}}$ synthetic dataset. We keep the other model parameters $\{b_\delta, \tilde{\sigma}_\epsilon\}$ free as well, whose posteriors are shown on Fig. 3.2 (the $\Lambda = 0.1 h \text{ Mpc}^{-1}$ contours). The gray band shows the modes with $k > \Lambda$, while the color bar traces the sample number n . The top panel shows the ratio of the sampled \hat{s} power spectra and the corresponding ground-truth \hat{s}_{true} power spectrum. Clearly, $P_{\hat{s}_{\text{true}}}$ is well within the posterior. The middle panel shows the binned residuals, $\Delta_{\hat{s}} = \hat{s} - \hat{s}_{\text{true}}$, which are found to be centered around 0 as expected. Note that, above the cutoff, we have $\Delta_{\hat{s}} \rightarrow -\langle \hat{s}_{\text{true}} \rangle$, since for these modes, the sampled \hat{s} values follow the prior which is zero-centered. In addition, we also show the 68% CL intervals estimated from the Wiener-filter solution (68% CL^{lin.model}) and the prior (68% CL^{prior}). The bottom panel shows the power spectrum of the $\Delta_{\hat{s}}$ field. Alongside this, we also show the Wiener filter expectation of the posterior \hat{s} covariance as well as the prior. As can be seen, the modes below the cutoff agree well with the Wiener filter prediction, while above the cutoff of $\Lambda = 0.1 h \text{ Mpc}^{-1}$ they follow the prior covariance. As discussed above Eq. (3.23), the Wiener-filter solution is obtained by fixing b_δ and P_ϵ to their fiducial values. The corresponding trends are also found for other cutoff values Λ .

The yellow dotted line in the middle panel of Fig. 3.3 represents the residual between the Wiener-filter solution \hat{s}_{WF} , i.e. the linear model analytical prediction from above, and \hat{s}_{true} . Clearly, the sampled posterior is precisely centered around \hat{s}_{WF} , indicating that it is unbiased also in the case when b_δ and $\tilde{\sigma}_\epsilon$ are left free. Above the cutoff, indicated by the gray band, the sampled modes simply follow the prior, hence $\Delta_\delta \rightarrow -\langle \hat{s}_{\text{true}} \rangle \rightarrow 0$. However, due to the finite number of modes per k -bin, the calculated mean will not be zero exactly, but vary around it within the prior bounds, which is indeed what we see.

Finally, the bottom panel of Fig. 3.3 shows the power spectrum of the Δ_δ field. The theoretical expectation is that the power spectrum of Δ_δ in a given k -bin is Γ distributed, $P_{\Delta_\delta} \leftarrow \Gamma(a, b)$, where the shape parameter is $a = N_{\text{mode}}/2$, N_{mode} being the number of modes within the Fourier space shell centered on k , and the scale parameter is $b = 2C_{\text{WF}}(k)$. This conclusion follows from considering the distribution of a sum of squares of Gaussian-distributed variables, in this case \hat{s} . Since the mean of the Γ distribution is given by the product of its scale and shape parameter, it follows immediately that the expectation value of P_{Δ_δ} within a given k -bin is $C_{\text{WF}}(k)$.

This prediction is shown as the black line in the bottom panel of Fig. 3.3. We find good agreement with the sampled results, indicating that the posterior is fully explored by the sampler. We expect this to be the case, even though the Wiener filter calculation assumes fixed parameters, since the parameters $b_\delta, \tilde{\sigma}_\epsilon$ are very well constrained, so that the propagated effect of their variance is a subdominant contribution to the \hat{s} posterior variance. We also show the \hat{s} prior covariance for comparison, and as we can see, modes above the cutoff indeed follow the prior.

To further investigate the posterior, in Fig. 3.4, we plot, for three selected k -bins, both the histogram of Δ_δ and the corresponding Wiener filter prediction, $\hat{s}_{\text{WF}} - \hat{s}_{\text{true}}$. The latter is, of course, Gaussian and plotted as the dotted line. To guide the eyes, we also show a vertical line on zero, indicating the ground truth. The top panel of this figure depicts the inferred \hat{s} residual statistics of the linear forward model from Eq. (3.3). In this case, the predicted (Wiener-filter) and sampled posteriors fully agree. On the other hand, the bottom panel shows the same, but for the forward model including second-order bias (Eq. (3.5)). Here we see clear deviations between the analytical and sampled posteriors. That is, the \hat{s} posterior for this simple but nonlinear forward model is not well approximated by the Wiener-filter solution (see Section 3.5.1 and Appendix 3.8.2 for more details). This highlights the importance of going beyond the Wiener filter approach when trying to extract information from even mildly nonlinear scales.

Second-order bias

Next, we perform consistency tests of the EFT likelihood on the $\mathcal{D}_2^{\text{LINEAR}}$ datasets. These synthetic datasets are generated using the forward model in Eq. (3.5). That is, they additionally include a non-negligible quadratic bias contribution. Note that this contribution however still involves only the linearly evolved density field $\delta_{\Lambda_0}^{(1)}$. This results in a

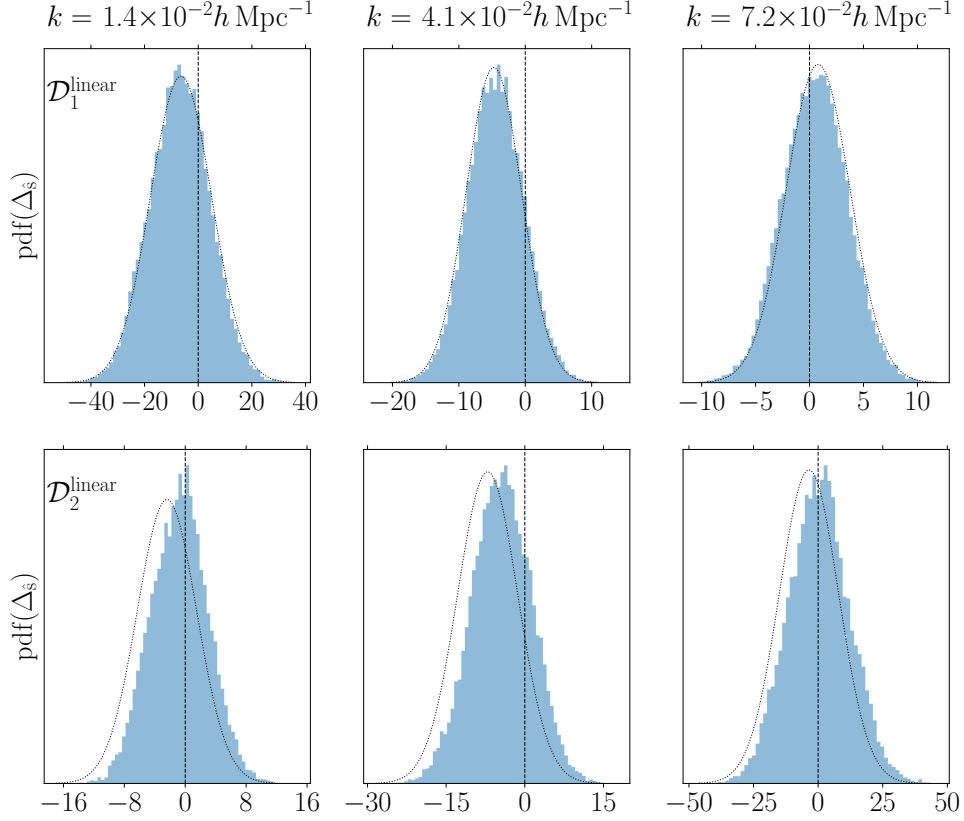


Figure 3.4: A closer look into the binned statistics of Δ_s . The units on the x -axis correspond to the units of the $\hat{s}(\mathbf{k})$ field, which is dimensionless based on our discrete Fourier convention. The k -bin centers are indicated at the top and are the same for both figures in each column. The dotted line indicates the predicted Wiener filter posterior for the statistics of the initial conditions, while the histogram represents the posterior obtained through sampling. For the top row, the setup is identical to that of Fig. 3.3, i.e. it shows the posterior of the statistics of Δ_s for the case of forward model from Eq. (3.3) applied to the $\mathcal{D}_1^{\text{linear}}$ dataset. As can be seen, the sampled posterior follows the Wiener filter prediction closely for all k bins. The bottom panel shows the corresponding results when applying the forward model with second-order bias, from Eq. (3.5), to $\mathcal{D}_2^{\text{linear}}$. Here the deviation from the Gaussian posterior of the Wiener filter prediction is prominent, and the latter is generally more biased than the sampled posterior.

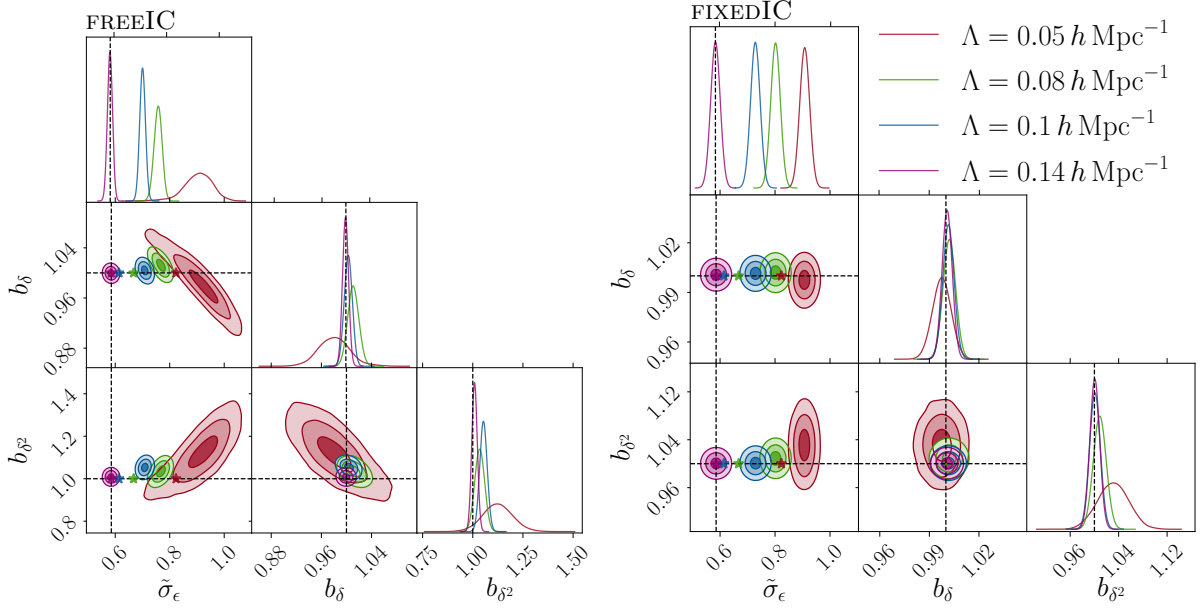


Figure 3.5: Same as Fig. 3.2 but for the forward model from Eq. (3.5). Again, the left panel shows the FREEIC case, while the right one shows the FIXEDIC case. Good agreement is again found between the means of the two posteriors. Note the difference in the inferred parameter ranges between the FIXEDIC and FREEIC posteriors. The dashed lines indicate the expected values for $\{b_\delta, b_{\delta^2}\}$ parameters, as well as the fiducial value for the noise level parameter $\tilde{\sigma}_\epsilon$ used for generating the synthetic data. Note that the inferred noise level increases towards smaller cutoffs for both the FIXEDIC and FREEIC posteriors. We further elaborate on this in the text and indicate the predicted running of $\tilde{\sigma}_\epsilon$ parameter from Eq. (3.25) by the star symbols.

non-Gaussian posterior of the initial conditions, \hat{s} , about which nonetheless we are able to make some qualitative analytical statements (see Appendix 3.8.2 and Appendix 3.9).

First, we analyze the posterior of the inferred parameters $\{b_\delta, b_{\delta^2}, \tilde{\sigma}_\epsilon\}$ shown in Fig. 3.5. As before, we consider forward models with different cutoffs Λ . Here, since the synthetic data $\mathcal{D}_2^{\text{LINEAR}}$ is generated with a nonzero b_{δ^2} , it introduces mode couplings across the whole available range of modes up to the synthetic data cutoff of $\Lambda_0 = 0.14 h \text{ Mpc}^{-1}$.

We first look at the noise amplitude $\tilde{\sigma}_\epsilon$. Fig. 3.5 shows that the inferred value is a function of the cutoff for both FIXEDIC and FREEIC cases; the inferred value is largest for the forward model with $\Lambda = 0.05 h \text{ Mpc}^{-1}$ and lowest for $\Lambda = \Lambda_0 = 0.14 h \text{ Mpc}^{-1}$. This can be understood as follows. The synthetic dataset $\mathcal{D}_2^{\text{LINEAR}}$ is generated using the forward model from Eq. (3.5), but with a cutoff of $\Lambda_0 = 0.14 h \text{ Mpc}^{-1}$ (see also Tab. 3.1). We can then split the linear density field from which the synthetic dataset is constructed

as

$$\delta_{\Lambda_0}^{(1)}(\mathbf{x}) = \delta_{\Lambda}^{(1)}(\mathbf{x}) + \delta_s^{(1)}(\mathbf{x}), \quad (3.24)$$

where $\delta_{\Lambda}^{(1)}$ and $\delta_s^{(1)}$ represent the parts of $\delta_{\Lambda_0}^{(1)}$ containing modes up to Λ , and from Λ to Λ_0 , respectively. Thus, $\delta_{d,\Lambda}$ contains a contribution $b_{\delta^2}(\delta_s^{(1)})^2(\mathbf{x})$ (see Eq. (3.5)). Since $\delta_s^{(1)}$ is uncorrelated with $\delta_{\Lambda}^{(1)}$, this contribution to the data corresponds to an additional noise that is absorbed in P_{ϵ} in the inference. We thus expect P_{ϵ} to shift by the power spectrum of $(\delta_s^{(1)})^2$, leading to

$$P_{\epsilon}(k) = P_{\epsilon}^{\text{no } b_{\delta^2}} + 2b_{\delta^2} \int_{\mathbf{p}} P_{\text{L}}^{[\Lambda, \Lambda_0]}(p) P_{\text{L}}^{[\Lambda, \Lambda_0]}(|\mathbf{k} - \mathbf{p}|) \quad (k < \Lambda) \quad (3.25)$$

where $P_{\text{L}}^{[\Lambda, \Lambda_0]}(p) = W_{\Lambda_0}(p)[1 - W_{\Lambda}(p)]P_{\text{L}}(p)$

and $P_{\text{L}}(p)$ is the linear power spectrum. Notice that only Fourier modes in the shell $[\Lambda, \Lambda_0]$ contribute. Evaluating this $P_{\epsilon}(k)$ at $k = \Lambda$ leads to the results represented with stars in Fig. 3.5. Note that for the case of $\Lambda = \Lambda_0$ (purple) there is no running of $\tilde{\sigma}_{\epsilon}$ parameter and the corresponding star is right in the center of the posterior contours for this case. In general, we find that the analytical result predicts the right trend, although the shift in the sampled posterior mean is generally larger than the prediction, in particular for Λ values that approach Λ_0 . The most likely explanation is that the inferred $\tilde{\sigma}_{\epsilon}$ also has to absorb the scale-dependence of the induced noise, since Eq. (3.25) has a significant k -dependence in particular if Λ is not much smaller than Λ_0 .

We now turn to the bias parameters. First, we expect no running of b_{δ} with the cutoff Λ , as the former only multiplies the linear density $\delta_{\Lambda}^{(1)}$ in the forward model Eq. (3.5). In other words, $b_{\delta,\Lambda} = b_{\delta,\Lambda_0}$ for all Λ . In fact, this argument can be made rigorous by examining the maximum likelihood point of the EFT likelihood (see for example Section 4 of [102]). Note that the maximum likelihood argument assumes the initial conditions \hat{s} fixed to the ground truth \hat{s}_{true} . That is, strictly speaking, the argument applies only to the FIXEDIC case. However, we generally expect the FREEIC posterior to overlap the FIXEDIC one, and Fig. 3.5 confirms that this is indeed the case.

As for b_{δ^2} , in Appendix 3.9 we derive the running using a similar approach and show that it vanishes as well; that is, $b_{\delta^2,\Lambda} = b_{\delta^2,\Lambda_0}$. We plot the results for b_{δ^2} for the different FREEIC chains in Fig. 3.6. There is a residual shift away from the expected value of b_{δ^2} when considering synthetic datasets with our fiducial noise level ($\tilde{\sigma}_{\epsilon} = 0.586$), as shown by the yellow data points. While we expect the model from Eq. (3.5) to be more accurate toward lower Λ , we in fact observe that the shift in $b_{\delta^2,\Lambda}$ with respect to the expected result increases as we lower the cutoff. The most plausible explanation for this is a prior volume effect resulting from the weakening constraint on b_{δ^2} and the growing degeneracy with $\tilde{\sigma}_{\epsilon}$ toward lower Λ (see Fig. 3.5). To confirm this, we performed an inference with substantially lower noise ($\tilde{\sigma}_{\epsilon} = 0.002$), indicated by the red points in Fig. 3.6. Indeed, the systematic shift is substantially reduced, showing that more informative datasets help

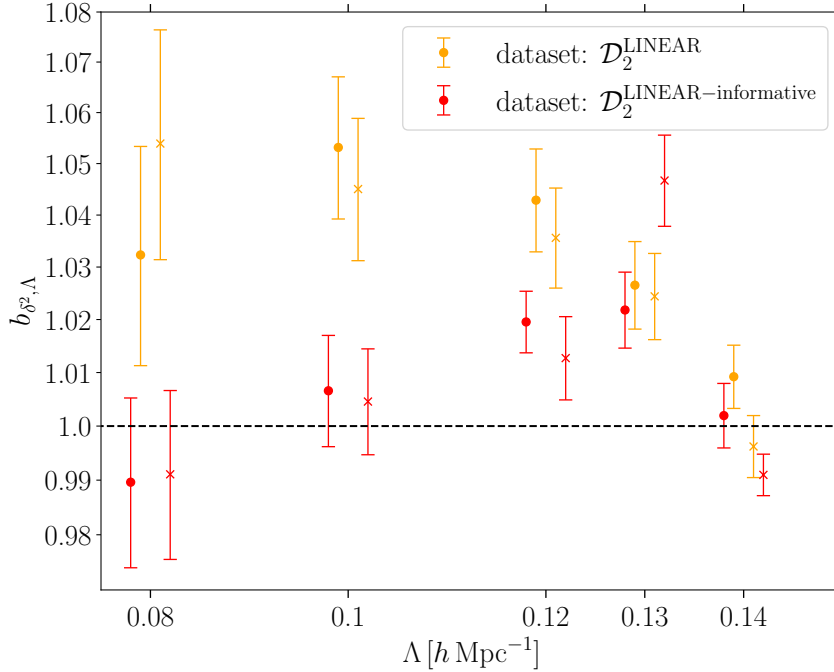


Figure 3.6: Mean a posteriori second-order bias parameter for the forward model from Eq. (3.5) inferred from different realizations (indicated by circle and cross) of the synthetic data $\mathcal{D}_2^{\text{LINEAR}}$ (yellow). The red color represents the inferred values of b_{δ^2} on $\mathcal{D}_2^{\text{LINEAR-informative}}$ dataset. The error bars indicate 68% CL intervals estimated from the chains.

with breaking the degeneracy. To conclude, the expected values for the bias parameters $\{b_\delta, b_{\delta^2}\}$ are precisely recovered at 68% – 95% CL by our FREEIC posteriors. The only stronger deviation occurs for forward models with cutoffs $0.1 h \text{ Mpc}^{-1} < \Lambda < \Lambda_0$, which persists even in the case of highly informative data (see Fig. 3.6). It is possible that including the subleading, k^2 contribution to the noise, which is expected to become more important as Λ approaches Λ_0 , would help with this residual shift.

Next, we focus on analyzing the \hat{s} posterior. As before, we look at the first and second moments of the posterior of initial conditions. However, now the Wiener filter prediction for the posterior mean is less accurate than for the case shown in Fig. 3.3 due to the presence of the quadratic term in the forward model. Fig. 3.7 compares the estimated 68% CL from the chain samples (dotted lines, with bands indicating sample variance) with the Wiener-filter expectation for the variance (dashed lines). As expected, the deviation from the Wiener-filter solution is stronger, at fixed k/Λ , as we go toward higher cutoffs. This is both because the typical amplitude of density fluctuations grows

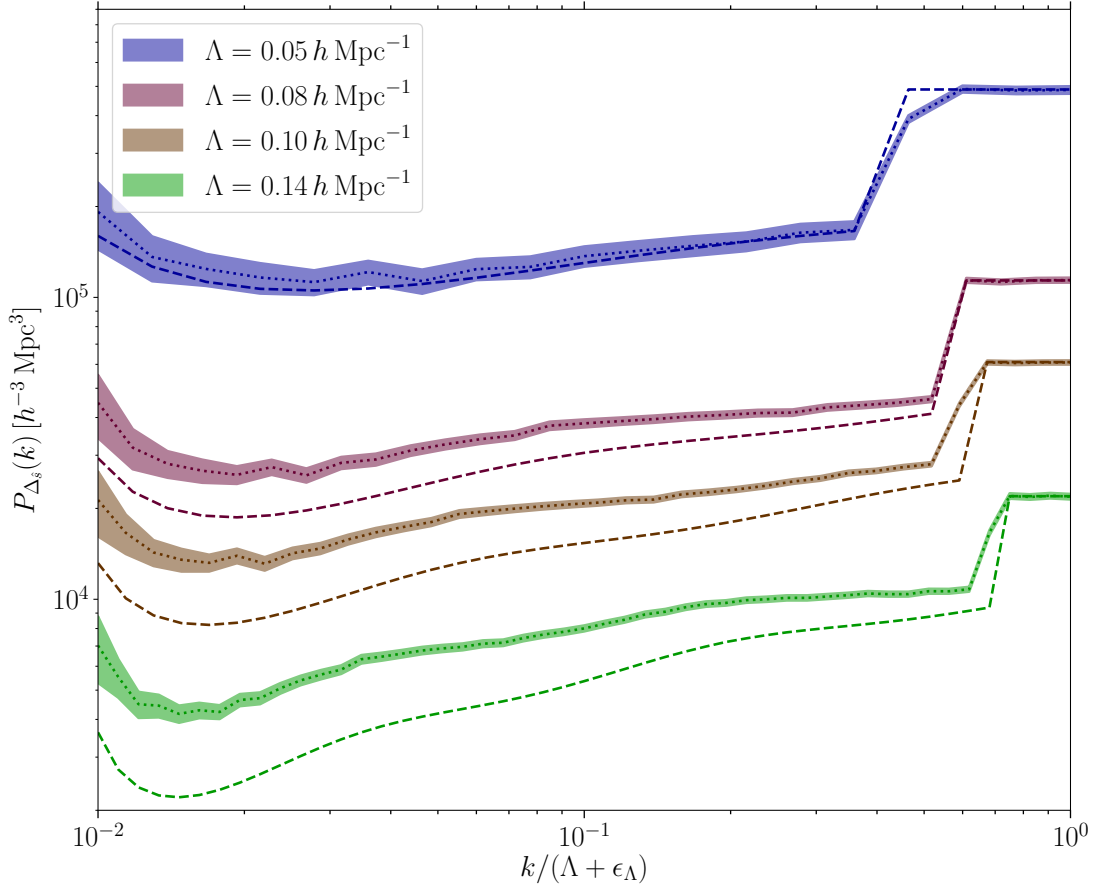


Figure 3.7: The power spectrum of Δ_s for MCMC chains using the forward model from Eq. (3.5) for different cutoffs Λ , shown as a function of $k/(\Lambda + \epsilon_\Lambda)$. We rescale the x -axis for plotting purposes, such that all the modes can be seen for all forward models with different cutoffs, suitably adjusting the ϵ_Λ parameter for each cutoff Λ . The inference here was performed on $\mathcal{D}_{2,b}^{\text{LINEAR}}$ (inferences on the $\mathcal{D}_{2,a}^{\text{LINEAR}}$ dataset yield consistent results). The shaded regions represent the mean and 68% CL of the sampled posterior. In addition, we also show the Wiener filter prediction, assuming a fully linear forward model. This approximation deteriorates toward higher cutoffs at fixed k/Λ , both because the mode coupling in the forward model becomes more important, and because \hat{s} is better constrained (see also Appendix 3.8.2).

on smaller scales, and the uncertainty on \hat{s} shrinks.

We can in fact make some qualitative statements about the behavior seen in Fig. 3.7. In case of the quadratic bias forward model, the posterior of initial conditions contains terms proportional to $\sim \hat{s}^3$ and $\sim \hat{s}^4$, in addition to the terms $\sim \hat{s}$ and $\sim \hat{s}^2$ present

in the purely linear case. The \hat{s} posterior covariance depends on all of these terms, as demonstrated in Appendix 3.8.2. There, we consider under what conditions the posterior for \hat{s} can be approximated analytically and describe the cause for the discrepancy between the Wiener filter prediction and the sampled joint posterior.

3.5.2 1LPT and 2LPT forward models

We now turn toward the forward models involving nonlinear gravity, i.e. 1LPT and 2LPT. The forward models we consider in this section are the most realistic ones, and allow for reducing the degeneracy between the bias parameters and the scaling parameter, α , as we elaborate below. Furthermore, as already hinted in [102], these forward models promise information gains beyond the leading- and next-to-leading order power spectrum, and leading-order bispectrum. We leave it for future work to explicitly demonstrate this. Instead, below we focus on the performance of these forward models on synthetic datasets listed in Section 3.4.2.

Linearly biased case

The synthetic dataset used in this section is $\mathcal{D}_1^{2\text{LPT}}$, with different realizations denoted by $\mathcal{D}_{1,a}^{2\text{LPT}}$ and $\mathcal{D}_{1,b}^{2\text{LPT}}$. These correspond to a linearly biased tracer of the 2LPT-evolved matter density field, and we thus test the consequences of a mismatch in the nonlinear matter forward model. Further, owing to the latter, the exact degeneracy between b_δ and α that is present for trivial linear evolution is now broken, as $\delta_{n\text{LPT}}$ contains terms scaling $\propto \alpha$ and $\propto \alpha^2$ both multiplied by the same b_δ (see [104, 106] for more discussion).

Fig. 3.8 shows the parameter posteriors after explicitly marginalizing over the posterior of initial conditions. Here we compare the inferences employing the 1LPT (red and green contours) and 2LPT (blue and dark-purple contours) forward models. Evidently, the parameters α , b_δ , and $\tilde{\sigma}_\epsilon$, which are not expected to run, agree well among the two different gravity models. The running is not present since α is a cosmological parameter, while b_δ and $\tilde{\sigma}_\epsilon$ are protected from running thanks to the absence of nonlinear bias in the synthetic datasets $\mathcal{D}_1^{2\text{LPT}}$. Note however that $b_{\nabla^2\delta}$ is expected to absorb the effect of modes between Λ and Λ_0 , hence to be shifted from its fiducial value. We also observe an expected anti-correlation in the $b_\delta - \alpha$ plane, given that these two parameters appear together as a product in the linear bias term in the forward model.

Another interesting degeneracy is in the $b_{\nabla^2\delta} - \tilde{\sigma}_\epsilon$ plane, which shows positive correlation. This can be understood by recalling how these parameters affect the leading order observable, the power spectrum. The dominant contribution to the tracer power spectrum that contains $b_{\nabla^2\delta}$ is $-2k^2 b_{\nabla^2\delta} b_\delta P_L(k) \sim -k^{0.5}$ at $k \approx 0.1 h \text{ Mpc}^{-1}$, while the noise contribution scales as $\sim k^0$. Since these two contributions have a similar k dependence, but opposite signs, they result into a positive correlation between the two corresponding parameters.

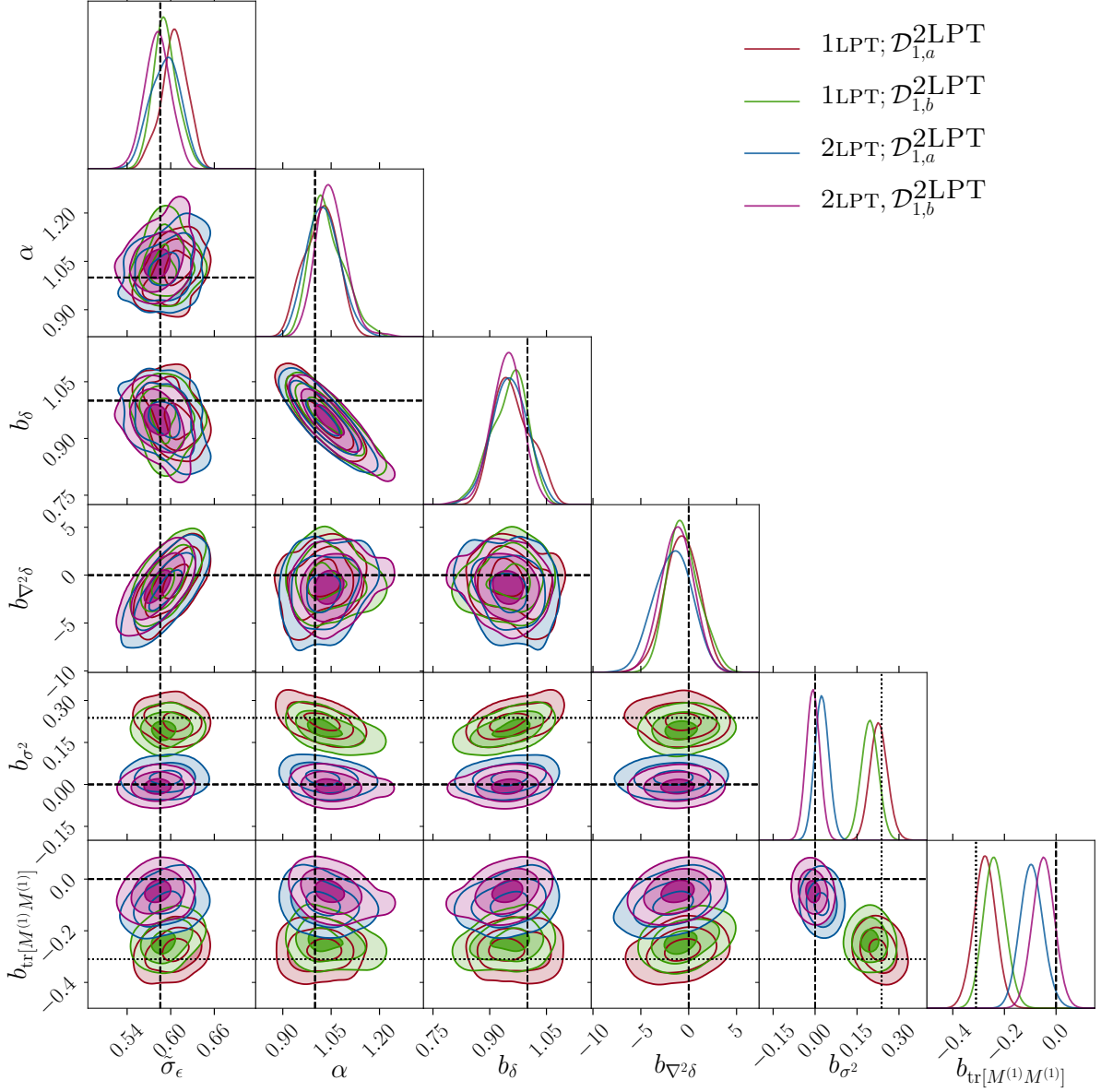


Figure 3.8: FREEIC parameter posteriors for the 2LPT and 1LPT inferences. The synthetic datasets used for this inference are $\mathcal{D}_{1,a}^{2\text{LPT}}$ and $\mathcal{D}_{1,b}^{2\text{LPT}}$, both generated at $\Lambda_0 = 0.14 h \text{Mpc}^{-1}$. The dashed lines indicate the fiducial input parameters of the synthetic datasets. The blue and dark-purple contours show 2LPT inferences, while the red and green contours show the same for the 1LPT inferences, each for the two independent datasets. The cutoff used for all the forward models is $\Lambda = 0.1 h \text{Mpc}^{-1}$. Note the positive correlation in the $b_{\nabla^2 \delta} - \bar{\sigma}_\epsilon$ plane and negative correlation in the $b_\delta - \alpha$ plane, as well as the shifts away from zero in the higher-order bias parameters b_{σ^2} and $b_{\text{tr}[M^{(1)}M^{(1)}]}$ in cases of the 1LPT inferences. The dotted lines indicate the expected values of b_{σ^2} and $b_{\text{tr}[M^{(1)}M^{(1)}]}$ from Lagrangian perturbation theory (see Eq. (3.26) and Appendix 3.11 for more details).

Another interesting feature can be seen when comparing the two different forward models in the bottom two rows of Fig. 3.8. Namely, the higher-order bias coefficients inferred using the 1LPT forward model are shifted away from their fiducial values of zero. The shifts of these bias coefficients can in fact be predicted using a second-order LPT calculation. Specifically, by solving for the displacement field and then substituting that back into the second-order bias expansion, one can derive the relations between the bias coefficients in the 1LPT and 2LPT forward model. Following the calculation done in Appendix 3.11, one derives the following relations between the bias coefficients of the two forward models

$$\begin{aligned} b_{\delta}^{1\text{LPT}} &= b_{\delta}^{2\text{LPT}}, \\ b_{\sigma^2}^{1\text{LPT}} &= b_{\sigma^2}^{2\text{LPT}} + \frac{3}{14}, \\ b_{\text{tr}[M^{(1)}M^{(1)}]}^{1\text{LPT}} &= b_{\text{tr}[M^{(1)}M^{(1)}]}^{2\text{LPT}} - \frac{3}{14}. \end{aligned} \quad (3.26)$$

These values are indicated with dotted lines in Fig. 3.8, and are within 68–95%CL of the corresponding 1LPT posteriors. The results we have discussed so far were obtained using the likelihood from Eq. (3.15), i.e. the unmarginalized likelihood. Using the marginalized likelihood from Eq. (3.16) gives entirely consistent results (see Appendix 3.12). However, the marginalized likelihood offers the important advantage of a reduced correlation length in the remaining parameters $\alpha, \tilde{\sigma}_\epsilon$. Namely, marginalizing over the bias parameters allows for a $\sim 60\%$ reduction in the correlation length of the α parameter (see the top panel of Fig. 3.9).⁴

This in turn means that for the same CPU time, the number of effective samples produced by the marginalized likelihood is correspondingly increased by a factor of 1.6. This improvement is expected to become more significant with increasing number of marginalized bias parameters, and suggests that the marginalized likelihood should be preferred, especially for higher-order bias models.

As a final remark, we also show the posterior of initial conditions within different k -bins in Fig. 3.10. As anticipated based on the previously shown results in Fig. 3.4, the posterior is indeed non-Gaussian, showing stronger deviations from the Gaussian case as one goes toward smaller scales (reflected in the heavier tails of the distribution). Furthermore, even on the largest scales covered by our simulated volume, the prediction from the Wiener filter is biased with respect to the inferred posterior which is correctly centered around the ground truth (see the left-most panel).

⁴Note that the correlation length was estimated by taking the average over three (two) independent chains for the unmarginalized (marginalized) likelihoods, respectively.

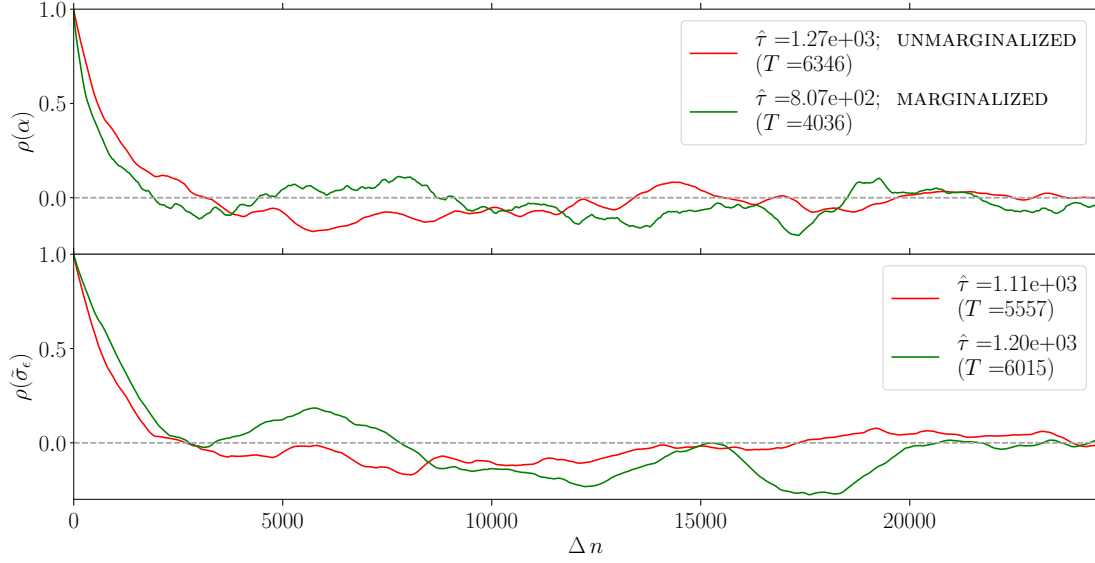


Figure 3.9: The normalized auto-correlation function (see Eq. (3.47)) for the 2LPT forward model applied to the $\mathcal{D}_1^{2\text{LPT}}$ dataset for α (top panel) and $\tilde{\sigma}_\epsilon$ (bottom panel). The x -axis shows the separation between samples in the chain, denoted with Δn . The labels in the legend show the estimated correlation length $\hat{\tau}$ and T , the corresponding maximal sample separation considered for making this estimate (see Appendix 3.10 for more details). In both cases, we compare the chains using the unmarginalized (red) and marginalized (green) likelihoods (see Eqs. (3.15)–(3.16)). A faster decay of the auto-correlation function can be seen for the marginalized likelihood in case of α , while for $\tilde{\sigma}_\epsilon$ the correlation lengths are comparable.

Biased tracers

The synthetic datasets from Section 3.5.2 consisted merely of the evolved 2LPT matter field, rescaled by the linear bias b_δ with added Gaussian noise. Here, we consider synthetic datasets with nonzero higher-order bias coefficients, as well as a cutoff mismatch. This means that we test for the ability of our forward model to extract correct α values from the biased tracers while marginalizing over plausible initial conditions realizations. As before, the $\mathcal{D}_2^{2\text{LPT}}$ dataset is generated using a cutoff $\Lambda = 0.14 h \text{ Mpc}^{-1}$, i.e. restricting to mildly nonlinear scales (recall that all synthetic data sets are at $z = 0$). A more realistic test case would adopt dark matter halo or simulated galaxy fields identified in cosmological N-body or hydrodynamic simulations as input. We leave this for future work.

In case of the inference performed on $\mathcal{D}_2^{2\text{LPT}}$, we currently do not have analytical predictions for the cutoff dependence of the bias coefficients; once these become available,

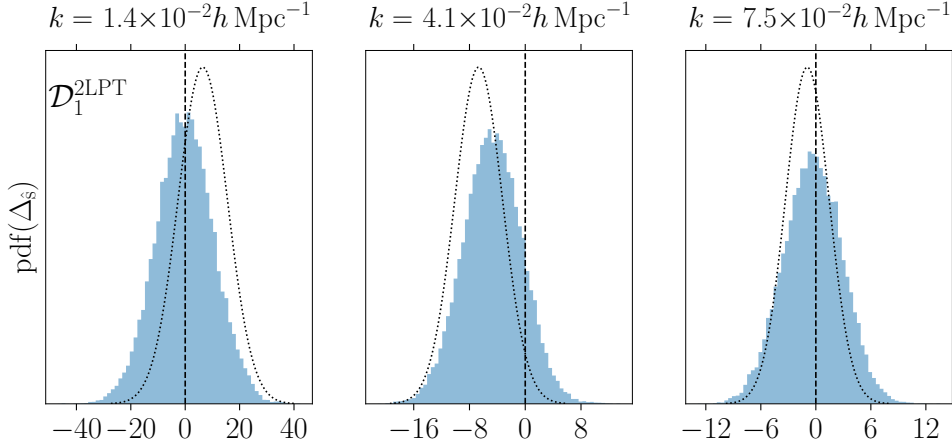


Figure 3.10: Same as Fig. 3.4 but for the 2LPT forward model from Eq. (3.11) applied to the $\mathcal{D}_1^{2\text{LPT}}$ synthetic data. We can see that the inferred posterior differs significantly from the predicted Wiener-filter solution, and is closer to the ground truth, \hat{s}_{true} , as expected. Even on large scales, the deviation is significant. Note also that the distribution of \hat{s} becomes non-Gaussian, and increasingly so toward smaller scales.

we will be able to test the inference results on these bias parameters as well. Instead, we consider the value of α as a measure of the model performance, as this, being a cosmological parameter, should be consistent across different cutoffs Λ .

We focus here on the results using the unmarginalized likelihood from Eq. (3.15), shown in Fig. 3.11; when using the marginalized likelihood, we find consistent results, similar to what we discuss in Section 3.5.2. We find a difference between the fiducial noise level and the one inferred at a lower cutoff, similar to the case in Section 3.5.1. The reason is the same as there, namely the presence of second-order bias terms in the $\mathcal{D}_2^{2\text{LPT}}$ dataset and the fact that $\Lambda_0 > \Lambda$. Estimating this shift could be done similarly as in Eq. (3.25), but taking into account the presence of higher-order nonlinear terms. We leave this calculation for future work.

Turning to the α parameter, both the $\Lambda = 0.14 h \text{ Mpc}^{-1}$ (blue) and $\Lambda = 0.1 h \text{ Mpc}^{-1}$ (red) FIXEDIC posteriors consistently infer the correct value as expected. For the FREEIC case, the $\Lambda = 0.14 h \text{ Mpc}^{-1}$ forward model posterior (purple contours) is able to recover the fiducial α value, and is furthermore consistent with the corresponding FIXEDIC posterior (blue contours). However, the $\Lambda = 0.1 h \text{ Mpc}^{-1}$ FREEIC posterior (green contours) shows a preference for smaller α , and excludes the fiducial α value at 95%CL contour. Similar systematic shifts can also be observed in the bias parameters.

Given that we have already found evidence for prior-volume effects in the case of a linear gravity model with quadratic bias expansion in Section 3.5.1 (see Fig. 3.6) it is natural to suspect a similar cause here. However, this discrepancy could also point toward

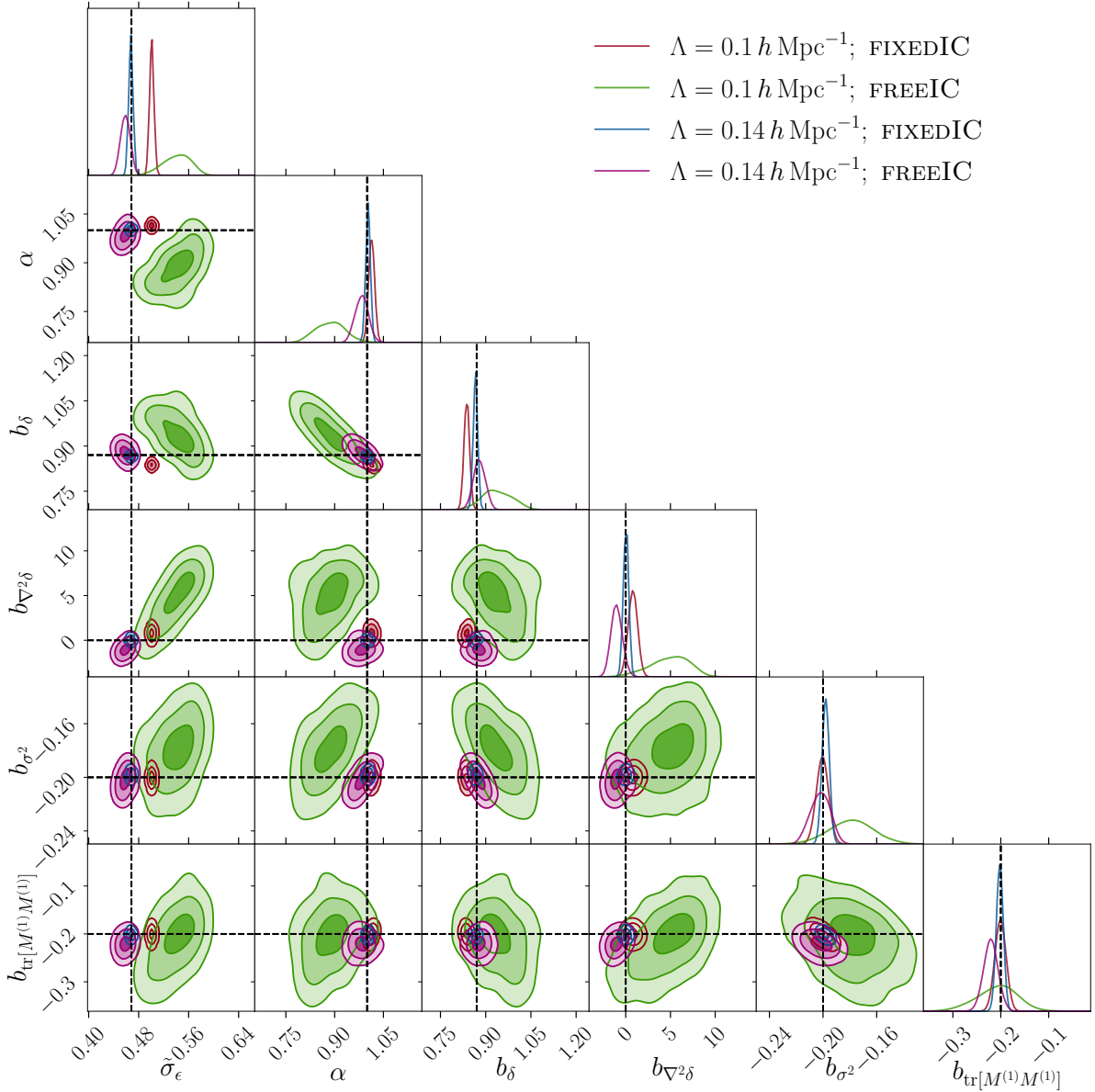


Figure 3.11: Parameter posteriors for the $\mathcal{D}_2^{2\text{LPT}}$ synthetic dataset using $\Lambda = 0.1 h \text{ Mpc}^{-1}$ (green and red contours) and $\Lambda = 0.14 h \text{ Mpc}^{-1}$ forward models (blue and purple contours) with 2LPT displacement fields. We contrast the FIXEDIC and the FREEIC inference chains for both forward models. As expected, the $\Lambda = 0.14 h \text{ Mpc}^{-1}$ FREEIC forward model recovers consistent results as the corresponding FIXEDIC case. However, the $\Lambda = 0.1 h \text{ Mpc}^{-1}$ FREEIC chains, i.e. those with a cutoff mismatch, show a shift with respect to the FIXEDIC contour. Possible explanations for this are discussed in the text.

the necessity of higher-order bias terms, or a scale-dependent or density-dependent noise covariance. We leave the investigation of the root cause for upcoming work.

3.6 Conclusions and Summary

In this chapter, we have investigated the robustness of field-level inference based on the EFT framework with respect to mismatch between theory and data, as well as the ability to constrain the amplitude of initial conditions (σ_8) when marginalizing over the initial conditions. Such tests do not only test the robustness of the chosen forward model and likelihood, but also allow for a better understanding regarding what types of forward model physics is necessary in order to capture all relevant effects and to obtain unbiased inference of the cosmological parameters.

We have focused on several different types of forward models described in Section 3.2, each probing different limits of our general forward model from Eq. (3.13), as well as different types of likelihood, one with explicit marginalization over bias parameters (Eq. (3.16)) and the other without (Eq. (3.15)). We perform all the tests on a suite of synthetic datasets with realistic noise levels, described in Section 3.4.

We have demonstrated in Section 3.5.1 that, in the case of the purely linear forward model (see Eq. (3.3)), our sampling approach coupled with the EFT likelihood attains a full exploration of the high-dimensional posterior, which in this case can be derived analytically (see Fig. 3.2–3.3 as well as the top panel of Fig. 3.4). This is a nontrivial result given the high-dimensional ($N_{\text{dim}} \sim 10^5 - 10^6$) posterior surface involved.

In Section 3.5.1 we have considered a simple but nontrivial extension by adding the b_{δ^2} term in the bias expansion as given in Eq. (3.5). This term leads to a mode coupling between two linear density fields, which in turn yields a non-Gaussian posterior of the initial conditions, as can be seen from the bottom panel of Fig. 3.4 as well as Fig. 3.7. Nevertheless, the inferred parameters show the expected behavior also in this non-Gaussian case (Fig. 3.5). We further find good agreements between the FIXEDIC, where the initial conditions are fixed to their ground-truth values, and FREEIC posteriors. This illustrates that our forward modeling and sampling approaches explore the posterior around the correct solution.

We have also examined cases which include a model mismatch in the 2LPT gravity model, by way of choosing a lower cutoff in the inference than the one used to generate the synthetic data. As synthetic datasets, we consider matter fields in Section 3.5.2 and nonlinearly biased tracer fields in Section 3.5.2, both including white Gaussian noise. For these cases, we find that, even in the presence of model mismatch, the EFT likelihood is still able to obtain unbiased estimates of cosmological parameters, specifically α which is our proxy for σ_8 . We have also demonstrated the advantage of using the marginalized over the unmarginalized likelihood from Eqs. (3.16)–(3.15) respectively, by showing a significantly reduced correlation length, in particular for α .

We do find signs of a mild discrepancy in the inferred α value in the case for the synthetic data set including nonlinear bias, when also allowing for a Λ mismatch. We leave the exploration of possible causes to upcoming work; while prior volume effects could be responsible, higher-order bias terms and density-dependent noise could also be relevant for this particular data set. The flexibility in generating different synthetic data sets will allow for a disentangling of the possible causes. Such future tests should also include the generalization to synthetic data involving nontrivial noise, such as scale- and density-dependent or non-Gaussian (for example, Poisson) noise.

3.7 Appendix A: Fourier space convention

Below, we summarize the Fourier convention and notation we adopt throughout this chapter. In particular, we give the explicit relation between the Fourier- and Hartley-representation of \hat{s} . The latter is of relevance for understanding our prior choices in Section 3.3.1 and following the calculations done in Appendix 3.8.1 and Appendix 3.8.2.

First, we define the forward Fourier transform $f(\mathbf{k}) \equiv \{\mathbb{F}f(\mathbf{x})\}$ and its inverse transform $f(\mathbf{x}) \equiv \{\mathbb{F}^{-1}f(\mathbf{k})\}$ as

$$\begin{aligned} f(\mathbf{k}) &\equiv \int d^3\mathbf{x} f(\mathbf{x}) e^{-i\mathbf{k}\cdot\mathbf{x}} \equiv \int_{\mathbf{x}} f(\mathbf{x}) e^{-i\mathbf{k}\cdot\mathbf{x}}, \\ f(\mathbf{x}) &\equiv \int \frac{d^3\mathbf{k}}{(2\pi)^3} f(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}} \equiv \int_{\mathbf{k}} f(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}}. \end{aligned}$$

In practice, we operate on finite grids, thus we use the discrete Fourier transforms given by

$$\begin{aligned} \delta(\mathbf{k}) &= \sum_i^{N_g^3} \delta(\mathbf{x}_i) e^{-i\mathbf{k}\cdot\mathbf{x}_i}, \\ \delta(\mathbf{x}) &= \frac{1}{N_g^3} \sum_{\mathbf{k}_i} \delta(\mathbf{k}_i) e^{i\mathbf{k}_i\cdot\mathbf{x}}, \end{aligned}$$

where $\mathbf{k} \in (n_x, n_y, n_z)k_F$ with $k_F = 2\pi/L$, and $n_i \in \{-N_g/2, \dots, N_g/2\}$. The Nyquist frequency is given by $k_{\text{Ny}} \equiv N_g k_F/2$. L stands for the box size. With this, the two-point prior of the initial conditions \hat{s} becomes

$$\langle \hat{s}(\mathbf{n}k_F) \hat{s}(\mathbf{n}'k_F) \rangle = \frac{1}{L^3} \delta_D^{\mathbf{n}, -\mathbf{n}'} P_{\hat{s}, \hat{s}}(\mathbf{n}k_F), \quad (3.27)$$

where $\delta_D^{\mathbf{n}, -\mathbf{n}'} = \delta_D^{n_x, -n'_x} \delta_D^{n_y, -n'_y} \delta_D^{n_z, -n'_z}$, with $\delta_D^{n_i, n_j}$ representing the Kronecker delta. For clarity, we also write this Kronecker delta in wavenumber space as $\delta_D^{\mathbf{k}, \mathbf{k}'}$. For a field $\hat{s}(\mathbf{x})$ drawn from a unit normal distribution in real space, it follows that $P_{\hat{s}, \hat{s}} = L^3 N_g^3$, and hence $\langle \hat{s}(\mathbf{n}k_F) \hat{s}(\mathbf{n}'k_F) \rangle = N_g^3 \delta_D^{\mathbf{n}, -\mathbf{n}'}$.

In order to implement the cutoff Λ , we use the isotropic sharp- k filter W_Λ defined as

$$W_\Lambda(\mathbf{k}) = \Theta_{\text{H}}(k - \Lambda), \quad (3.28)$$

with Θ_{H} being the Heaviside function.

Finally, we note that any field $\hat{s}(\mathbf{k})$ can be represented either in the Fourier or Hartley convention, with our `LEFTfield` code utilizing the latter. The two representations are

related through (see [126] and Section 3.4 in [127] for more details)

$$\begin{aligned}\hat{s}(k) &\equiv \{\mathbb{H}\hat{s}\}(\mathbf{k}) = \text{Re}[\{\mathbb{F}\hat{s}\}(\mathbf{k})] - \text{Im}[\{\mathbb{F}\hat{s}\}(\mathbf{k})], \\ \hat{s}(k) &\equiv \{\mathbb{F}\hat{s}\}(\mathbf{k}) = \frac{\{\mathbb{H}\hat{s}\}(-\mathbf{k}) + \{\mathbb{H}\hat{s}\}(\mathbf{k})}{2} + i \frac{\{\mathbb{H}\hat{s}\}(-\mathbf{k}) - \{\mathbb{H}\hat{s}\}(\mathbf{k})}{2}.\end{aligned}\quad (3.29)$$

with \mathbb{F} and \mathbb{H} denoting Fourier and Hartley transforms respectively. This is the field whose mean, residuals, and variance is shown in the figures in Section 3.5.

3.8 Appendix B: Gaussian expectation for \hat{s} posterior

In this section, we present efforts towards analytical understanding of the shapes of \hat{s} posteriors for the forward models represented by Eq. (3.3) (Appendix 3.8.1) and Eq. (3.5) (Appendix 3.8.2), restricting to the case where the bias parameters and noise amplitude are fixed to the ground truth. It is much more difficult to obtain an analytical expression for the posterior when also varying the latter.

In the linear case, an analytical form of the posterior exists, whose mean and variance coincides with the Wiener-filter solution (see, for example [34, 128] and references therein). In the nonlinear case however, only a perturbative approach is possible and we elaborate on this in Appendix 3.8.2.

3.8.1 Appendix B.1: Linear model

As discussed around Eq. (3.3), the covariance structure of cosmological initial conditions is diagonal in Fourier space. Specifically, using the Fourier-space representation of our prior covariance (see our Fourier convention from Appendix 3.7), one obtains

$$S_{\mathbf{k}'}^{\mathbf{k}} = N_g^3 \delta_D^{\mathbf{k},\mathbf{k}'},$$

where $\delta_D^{\mathbf{k},\mathbf{k}'}$ represents the Kronecker delta. The noise is likewise assumed to be Gaussian with diagonal covariance $(C_\epsilon)_{\mathbf{k}'}^{\mathbf{k}}$ related to P_ϵ (see Eq. (3.14)) as

$$(C_\epsilon)_{\mathbf{k}'}^{\mathbf{k}} = \delta_D^{\mathbf{k},\mathbf{k}'} P_\epsilon.$$

These two assumptions allow us to derive the expected posterior on \hat{s} . In order to more easily see this, we can rephrase Eq. (3.3) as follows

$$\begin{aligned}\delta_d^{\mathbf{k}} &= R_{\mathbf{k}'}^{\mathbf{k}} \hat{s}^{\mathbf{k}'} + \epsilon^{\mathbf{k}}, \\ R_{\mathbf{k}'}^{\mathbf{k}} &= \delta_D^{\mathbf{k},\mathbf{k}'} b_\delta T(\mathbf{k}),\end{aligned}$$

where repeated indices are summed over; in the following, we will drop the repeated indices. We also drop the explicit α dependence, since here we are only interested in the

posterior of initial conditions with α fixed to the ground truth. In the following, we will further fix the parameters b_δ and σ_ϵ ; only for this case can we derive the posterior for \hat{s} analytically.

The likelihood for δ_d can be derived by marginalizing over the noise distribution, which yields

$$\mathcal{P}(\delta_d|\hat{s}, b_\delta, \sigma_\epsilon) = \mathcal{G}(\delta_d; R\hat{s}, C_\epsilon),$$

with $R\hat{s}$ denoting the mean and C_ϵ denoting the covariance of this Gaussian. Going forward, we consider log-probabilities for convenience. This leads to (suppressing the conditional on $\{b_\delta, \sigma_\epsilon\}$ parameters for clarity)

$$\begin{aligned} -\ln \mathcal{P}(\delta_d, \hat{s}) &= -\ln \mathcal{P}(\delta_d|\hat{s}) - \ln \mathcal{P}(\hat{s}) \\ &= \frac{1}{2} (\delta_d - R\hat{s})^\dagger C_\epsilon^{-1} (\delta_d - R\hat{s}) + \frac{1}{2} \ln|2\pi C_\epsilon| + \frac{1}{2} \hat{s}^\dagger S^{-1} \hat{s} + \frac{1}{2} \ln|2\pi S| \\ &= \frac{1}{2} \left(\hat{s}^\dagger (R^\dagger C_\epsilon^{-1} R + S^{-1}) \hat{s} - \delta_d^\dagger C_\epsilon^{-1} R \hat{s} - \hat{s} R^\dagger C_\epsilon^{-1} \delta_d \right) \\ &\quad + \frac{1}{2} \left(\delta_d^\dagger C_\epsilon^{-1} \delta_d + \text{tr} \ln C_\epsilon + \text{tr} \ln S + N_{\delta_d} \ln 2\pi + N_{\hat{s}} \ln 2\pi \right), \end{aligned} \quad (3.30)$$

where N_{δ_d} and $N_{\hat{s}}$ represent the total number of modes in the δ_d and \hat{s} fields, respectively. In our applications, these are always the same.

Defining $j = R^\dagger C_\epsilon^{-1} \delta_d$ and $(C_{\text{WF}})^{-1} = (S^{-1} + R^\dagger C_\epsilon^{-1} R)$ we can rewrite Eq. (3.30) as

$$\ln \mathcal{P}(\delta_d, \hat{s}) = \frac{1}{2} (\hat{s} - C_{\text{WF}} j)^\dagger (C_{\text{WF}})^{-1} (\hat{s} - C_{\text{WF}} j) + \text{const.} \quad (3.31)$$

where we have accumulated all the \hat{s} -independent terms inside const., i.e.

$$\text{const.} \equiv \frac{1}{2} \left(\delta_d^\dagger C_\epsilon^{-1} \delta_d + \text{tr} \ln C_\epsilon + \text{tr} \ln S + N_{\delta_d} \ln 2\pi + N_{\hat{s}} \ln 2\pi - j^\dagger C_{\text{WF}} j \right). \quad (3.32)$$

It is now clear that the posterior of \hat{s} is Gaussian:

$$\mathcal{P}(\hat{s}|\delta_d) = \mathcal{G}(\hat{s}; \hat{s}_{\text{WF}}, C_{\text{WF}}), \quad (3.33)$$

with mean \hat{s}_{WF} and covariance C_{WF} given as

$$\begin{aligned} \hat{s}_{\text{WF}} &= C_{\text{WF}} j \\ C_{\text{WF}} &= (1 + S R^\dagger C_\epsilon^{-1} R)^{-1} S. \end{aligned} \quad (3.34)$$

Substituting into the second line of Eq. (3.34) the expression for response R , noise covariance $C_\epsilon(k)$ and the prior S yields the following expression for the \hat{s} posterior covariance

$$(C_{\text{WF}})_{\mathbf{k}\mathbf{k}'} = \delta_D^{\mathbf{k},\mathbf{k}'} \left(1 + N_g^3 \frac{b_\delta^2 P_L(k)}{P_\epsilon} \right)^{-1} N_g^3. \quad (3.35)$$

For the results shown in the main text, we calculate Eq. (3.35) for every mode. We note that when comparing our analytical expression from Eq. (3.35) to the results for the Δ_{δ} power spectrum obtained from sampling shown in the bottom panel of Fig. 3.3 and in Fig. 3.7, we account for the fact that $P_{\Delta_{\delta}}$ is in fact Γ distributed within each k -bin. The shape parameter is given by $N_{\text{mode}}/2$, with N_{mode} being the number of modes within the Fourier space shell centered on k , while the scale parameter is $2C_{\text{WF}}(k)$ (see Section 3.5.1). We again emphasize that the posterior mean and variance Eq. (3.34) coincide with the Wiener filter result only for a linear forward model, Gaussian prior and likelihood, and fixed parameters $\alpha, b_{\delta}, \tilde{\sigma}_{\epsilon}$. In fact, Fig. 3.7 indicates that the Wiener filter estimate of the residual variance is biased low, i.e. C_{WF} is relatively lower than the actual variance $P_{\Delta_{\delta}}$, for nonlinear forward models.

3.8.2 Appendix B.2: Quadratic model

We now consider the quadratic bias model with linearized gravity (see Eq. (3.5)). Before continuing, we refer readers to Appendix 3.7 for our discrete Fourier convention. We start with writing out the full likelihood expression (dropping again the α dependence since this parameter is held fixed)

$$\ln \mathcal{L}(\delta_{d,\Lambda} | \delta_{\text{det},\Lambda}[\{O, b_O\}], \sigma_{\epsilon}) = -\frac{1}{2} \sum_{\mathbf{k} \neq 0}^{k_{\text{max}}} \left[\ln 2\pi\sigma_{\epsilon}^2 + \frac{1}{\sigma_{\epsilon}^2} |\delta_{d,\Lambda}(\mathbf{k}) - \delta_{\text{det},\Lambda}[\{O, b_O\}](\mathbf{k})|^2 \right],$$

which when specialized for the forward model from Eq. (3.5) reads

$$\begin{aligned} \ln \mathcal{L}(\delta_{d,\Lambda} | \delta_{\text{det},\Lambda}[\{O, b_O\}], \sigma_{\epsilon}) &= -\frac{1}{2} \sum_{\mathbf{k} \neq 0}^{k_{\text{max}}} \left[\ln 2\pi\sigma_{\epsilon}^2 + \frac{1}{\sigma_{\epsilon}^2} |\delta_{d,\Lambda}|^2(\mathbf{k}) \right. \\ &\quad - \frac{1}{\sigma_{\epsilon}^2} \delta_{d,\Lambda}^*(\mathbf{k}) \left((R\hat{s})(\mathbf{k}) + (R_2\hat{s}\hat{s})(\mathbf{k}) \right) \\ &\quad - \frac{1}{\sigma_{\epsilon}^2} \delta_{d,\Lambda}(\mathbf{k}) \left((R\hat{s})^*(\mathbf{k}) + (R_2\hat{s}\hat{s})^*(\mathbf{k}) \right) \\ &\quad \left. + \frac{1}{\sigma_{\epsilon}^2} \left| (R\hat{s})(\mathbf{k}) + (R_2\hat{s}\hat{s})(\mathbf{k}) \right|^2 \right], \end{aligned} \quad (3.36)$$

with R and R_2 operations defined as

$$\begin{aligned} R_{\mathbf{k}_1}^{\mathbf{k}} [\cdot]^{k_1} &\equiv \delta_D^{\mathbf{k}, \mathbf{k}_1} b_{\delta} W_{\Lambda}(\mathbf{k}_1) T(k_1) [\cdot]^{k_1} \\ R_2^{\mathbf{k}}_{\mathbf{k}_1, \mathbf{k}_2} [\cdot, \cdot]^{k_1, k_2} &\equiv b_{\delta^2} \frac{1}{N_g^3} \sum_{\mathbf{k}_1, \mathbf{k}_2} \delta_D^{\mathbf{k}, \mathbf{k}_1 + \mathbf{k}_2} W_{\Lambda}(\mathbf{k}_1) W_{\Lambda}(\mathbf{k}_2) T(k_1) T(k_2) [\cdot, \cdot]^{k_1, k_2}. \end{aligned} \quad (3.37)$$

Note that the R operator is the same as that in the linear forward model described in Appendix 3.8.1. The R_2 operator implements the second-order bias via a convolution in

Fourier space, with the kernel represented by the product of the two transfer functions and the sharp- k cutoffs.

We now add the Gaussian log-prior on \hat{s} to Eq. (3.36) and expand in powers of \hat{s} . This results in the following ordering of terms (repeated indices are summed over)

$$\begin{aligned}
o(\hat{s}) &: \frac{2}{\sigma_\epsilon^2} \delta_{d,\Lambda}^{\mathbf{k}} R_{\mathbf{k}_1}^{\mathbf{k}} \hat{s}^{\mathbf{k}_1} \\
o(\hat{s}^2) &: \left(\frac{1}{\sigma_\epsilon^2} R_{\mathbf{k}_1}^{\mathbf{k}} R_{\mathbf{k}_2}^{\mathbf{k}} - \frac{2}{\sigma_\epsilon^2} \delta_{d,\Lambda}^{\mathbf{k}} R_{2\mathbf{k}_1, \mathbf{k}_2}^{\mathbf{k}} + \delta_D^{\mathbf{k}_1, \mathbf{k}_2} N_g^{-3} \right) \hat{s}^{\mathbf{k}_1} \hat{s}^{\mathbf{k}_2} \\
o(\hat{s}^3) &: \frac{2}{\sigma_\epsilon^2} R_{\mathbf{k}_1}^{\mathbf{k}} R_{2\mathbf{k}_2, \mathbf{k}_3}^{\mathbf{k}} \hat{s}^{\mathbf{k}_1} \hat{s}^{\mathbf{k}_2} \hat{s}^{\mathbf{k}_3} \\
o(\hat{s}^4) &: \frac{1}{\sigma_\epsilon^2} R_{2\mathbf{k}_1, \mathbf{k}_2}^{\mathbf{k}} R_{2\mathbf{k}_3, \mathbf{k}_4}^{\mathbf{k}} \hat{s}^{\mathbf{k}_1} \hat{s}^{\mathbf{k}_2} \hat{s}^{\mathbf{k}_3} \hat{s}^{\mathbf{k}_4}.
\end{aligned} \tag{3.38}$$

In other words, the final log-posterior is given by (in matrix notation)

$$\begin{aligned}
\mathcal{H}(\hat{s}|\delta_{d,\Lambda}) \equiv -\ln \mathcal{P}(\hat{s}|\delta_{d,\Lambda}) &= \underbrace{2R^\dagger C_\epsilon^{-1} \delta_{d,\Lambda}}_{j^\dagger} \hat{s} + \frac{1}{2} \hat{s}^\dagger \underbrace{\left(R^\dagger C_\epsilon^{-1} R - 2R_2 C_\epsilon^{-1} \delta_{d,\Lambda} + \mathbb{1} N_g^{-3} \right)}_{(D')^{-1}} \hat{s} \\
&+ \frac{1}{2} \hat{s}^\dagger \underbrace{R^\dagger C_\epsilon^{-1} R_2^\dagger}_{\mathcal{M}^{(3)}} \hat{s} \hat{s} + \frac{1}{2} \hat{s}^\dagger \hat{s}^\dagger \underbrace{R_2^\dagger C_\epsilon^{-1} R_2}_{\mathcal{M}^{(4)}} \hat{s} \hat{s},
\end{aligned} \tag{3.39}$$

where we have introduced the third and fourth order coupling kernels with $\mathcal{M}^{(3)}$ and $\mathcal{M}^{(4)}$ respectively. Also, we have relabeled the operators in the quadratic and linear term with $(D')^{-1}$ and j respectively. In the absence of the b_{δ^2} term, the posterior covariance is given exactly by the Wiener-filter solution for posterior covariance, i.e. second line of Eq. (3.34). For the posterior given in Eq. (3.39), it is not straightforward to compute the corresponding first and second moments. Instead, we expand the posterior around the Wiener-filter solution. While this expansion is strictly only valid if the correction due to b_{δ^2} is small, this expansion nevertheless offers some interesting insights.

We thus define $\hat{s}' \equiv \hat{s} - \hat{s}_{\text{WF}}$, where \hat{s}_{WF} represents the Wiener filter prediction of the initial conditions given by Eq. (3.34). In this case, the formalism of Information Field Theory, as presented in [34], suggests the following diagrammatic representation of the solution (see also Section V.C of [34] and, for the Feynman rules, Section IV.A.2 of the same paper)

$$\begin{aligned}
\langle \hat{s}'(\hat{s}')^\dagger \rangle &= \text{---} + \text{---} \langle \text{---} + 2 \text{ perm.} + \text{---} + 5 \text{ perm.} \\
&+ \text{---} \text{---} + \text{---} \text{---} \text{---},
\end{aligned}$$

where the diagrams correspond to the following expressions

$$\begin{aligned}
k_1 \text{ --- } k_2 &= D'_{\mathbf{k}_1, \mathbf{k}_2}, \\
\text{---} \begin{array}{c} \diagup \\ \diagdown \end{array} &\sim (D')^{\mathbf{k}_1 \mathbf{k}'} (\mathcal{M}^{(3)})^{\mathbf{k}}_{lm} (D')^{ll_1} j_{l_1} (D')^{m \mathbf{k}_2}, \\
\begin{array}{c} \vdots \\ \text{---} \\ \vdots \end{array} &\sim (D')^{\mathbf{k}_1 \mathbf{k}} (\mathcal{M}^{(4)})^{\mathbf{k}}_{lmn} (D')^{ll_1} j_{l_1} (D')^{mm_1} j_{m_1} (D')^{n \mathbf{k}_2}, \\
\text{---} \bigcirc &\sim (D')^{\mathbf{k}_1 \mathbf{k}} (\mathcal{M}^{(4)})^{\mathbf{k}}_{lmn} (D')^{lm} (D')^{n \mathbf{k}_2}, \\
\bigcirc \text{---} \bigcirc &\sim (D')^{\mathbf{k} l} (\mathcal{M}^{(4)})^{\mathbf{k}}_{lmn} (D')^{mn},
\end{aligned} \tag{3.40}$$

where we assume that the external lines (without dots) are fixed at \mathbf{k}_1 and \mathbf{k}_2 as indicated in the first line.

We can see that at leading order the posterior covariance is given exactly by D' , while the corrections to it depend on the exact form of the coupling kernels $\mathcal{M}^{(3)}$ and $\mathcal{M}^{(4)}$. The evaluation of D' requires an explicit matrix inversion.

To avoid this, and keeping in mind that this expansion is only valid for small corrections to the Wiener-filter posterior, we expand D' to obtain the leading correction to the posterior covariance

$$D' = (C_{\text{WF}}^{-1} - 2R_2 C_\epsilon^{-1} \delta_{d,\Lambda})^{-1} \approx C_{\text{WF}} + 2C_{\text{WF}} (R_2 C_\epsilon^{-1} \delta_{d,\Lambda}) C_{\text{WF}} + \dots \tag{3.41}$$

Writing out the leading correction term, one obtains

$$[C_{\text{WF}} (R_2 C_\epsilon^{-1} \delta_{d,\Lambda}) C_{\text{WF}}]_{\mathbf{k}}^{\mathbf{k}'} \propto b_{\delta^2} (W_\Lambda T \delta_{d,\Lambda})(\mathbf{k} - \mathbf{k}'). \tag{3.42}$$

This implies that the leading correction to the covariance around the Wiener-filter solution only contributes to the off-diagonal elements. This is an expected result, given the structure of the mode coupling introduced by the $\sim (\delta_\Lambda^{(1)})^2$ term from Eq. (3.5). In order to compute the correction to the diagonal part of the posterior covariance shown in Fig. 3.7, one would need to compute the next-to-leading correction to the posterior covariance. At this order, one also has to include the shift of the maximum of the posterior from the Wiener-filter solution, which is also of order b_{δ^2} . This is a more involved calculation which we leave to future work.

The considerations above indicate that obtaining even approximate analytical posteriors for \hat{s} is very difficult already for the simplest nonlinear models. These difficulties are correspondingly exacerbated for more nonlinear models, such as those involving n LPT forward models. Thus, the explicit sampling approach appears to be the only path toward obtaining trustable posteriors for initial conditions inference using nonlinear forward models.

3.9 Appendix C: Running of b_{δ^2} with the cutoff

In this section, we describe the 1-loop calculation of the expectation value of the b_{δ^2} parameter as a function of the forward model cutoff Λ . In order to derive this relation, it is sufficient to look at the maximum likelihood point of the unmarginalized likelihood (see Eq. (3.15))

$$\frac{\partial}{\partial b_{\delta^2}} \ln \mathcal{L}(\delta_{d,\Lambda} | \delta_{\text{det},\Lambda}[\{O, b_O\}, \hat{s}], \sigma_\epsilon) = \sum_{\mathbf{k} \neq 0}^{\Lambda} \left[\frac{1}{\sigma_\epsilon^2} \left(\delta_\Lambda^{(1)} \right)^2(\mathbf{k}) (\delta_{d,\Lambda}(\mathbf{k}) - \delta_{\text{det},\Lambda}(\mathbf{k}))^* \right] = 0. \quad (3.43)$$

We have again suppressed the explicit α dependence within $\delta_{\text{det},\Lambda}$, since for the forward model from Eq. (3.5) we keep it fixed. Since σ_ϵ is a constant, we can factor it out and using the fact that the forward model here is given by Eq. (3.5) one can rearrange the above Eq. (3.43) to obtain

$$\sum_{\mathbf{k} \neq 0}^{\Lambda} \left[\left(\delta_\Lambda^{(1)} \right)^2(\mathbf{k}) \delta_{d,\Lambda}(-\mathbf{k}) \right] = \sum_{\mathbf{k} \neq 0}^{\Lambda} \left[b_{\delta,\Lambda} \left(\delta_\Lambda^{(1)} \right)^2(\mathbf{k}) \delta_\Lambda^{(1)}(-\mathbf{k}) + b_{\delta^2,\Lambda} \left(\delta_\Lambda^{(1)} \right)^2(\mathbf{k}) \left(\delta_\Lambda^{(1)} \right)^2(-\mathbf{k}) \right],$$

where we have explicitly stated the cutoff dependence of the bias coefficients, which holds in general and used the fact that the density fluctuation field is hermitian.

Before proceeding, we note that the above equation holds for a given realization of initial conditions, as explicitly stated in Eq. (3.43).

In what follows, we evaluate the MAP relation for b_{δ^2} at the ground-truth initial conditions \hat{s}_{true} . This is obviously the correct choice when comparing to FIXEDIC chains. However, since \hat{s}_{true} is expected to be in the typical set of the FREEIC posterior, the result can also be translated to FREEIC chains. After evaluating on the ground truth, we then take the ensemble average over data realizations. This allows us to compute the result analytically and gives the following (see also section 3 of [102])

$$b_{\delta^2,\Lambda_0} \sum_{\mathbf{k} \neq 0}^{\Lambda} \left\langle \left(\delta_\Lambda^{(1)} \right)^2(\mathbf{k}) \left(W_\Lambda \left(\delta_{\Lambda_0}^{(1)} \right)^2 \right)(-\mathbf{k}) \right\rangle = b_{\delta^2,\Lambda} \sum_{\mathbf{k} \neq 0}^{\Lambda} \left\langle \left(\delta_\Lambda^{(1)} \right)^2(\mathbf{k}) \left(\delta_\Lambda^{(1)} \right)^2(-\mathbf{k}) \right\rangle, \quad (3.44)$$

keeping only the non-zero correlators. Given that the procedure for evaluating all correlators is essentially the same, we focus only on the correlator from the left-hand side of Eq. (3.44). Evaluating, we get

$$\left\langle \left(\delta_\Lambda^{(1)} \right)^2(\mathbf{k}) \left(W_\Lambda \left(\delta_{\Lambda_0}^{(1)} \right)^2 \right)(-\mathbf{k}) \right\rangle = \int_{\mathbf{k}_1} \int_{\mathbf{k}_2} W_\Lambda(\mathbf{k} - \mathbf{k}_1) W_\Lambda(\mathbf{k}_1) W_{\Lambda_0}(\mathbf{k}_2 + \mathbf{k}) W_{\Lambda_0}(\mathbf{k}_2) \langle \delta^{(1)}(\mathbf{k} - \mathbf{k}_1) \delta^{(1)}(\mathbf{k}_1) \delta^{(1)}(\mathbf{k} - \mathbf{k}_2) \delta^{(1)}(\mathbf{k}_2) \rangle.$$

Using repeated Wick contractions one gets

$$\begin{aligned} \left\langle \left(\delta_{\Lambda}^{(1)} \right)^2(\mathbf{k}) \left(W_{\Lambda} \left(\delta_{\Lambda_0}^{(1)} \right)^2 \right)(-\mathbf{k}) \right\rangle &= 2 \int_{\mathbf{k}_1} W_{\Lambda}(\mathbf{k} - \mathbf{k}_1) W_{\Lambda}(\mathbf{k}_1) W_{\Lambda_0}(\mathbf{k}_1 - \mathbf{k}) W_{\Lambda_0}(-\mathbf{k}_1) \\ &\quad P_L(|\mathbf{k} - \mathbf{k}_1|) P_L(k_1) \\ &\quad + \delta_D(\mathbf{k}) \delta_D(-\mathbf{k}) \int_{\mathbf{k}_1} \int_{\mathbf{k}_2} W_{\Lambda}(\mathbf{k} - \mathbf{k}_1) W_{\Lambda}(\mathbf{k}_1) \\ &\quad W_{\Lambda_0}(\mathbf{k} - \mathbf{k}_2) W_{\Lambda_0}(\mathbf{k}_2) \\ &\quad P_L(k_1) P_L(k_2), \end{aligned} \quad (3.45)$$

with $P_L(k)$ representing the linear power spectrum. From the above equation, it is clear that the correlator is a diagonal matrix in Fourier space. Furthermore, using the fact that the sharp- k cutoff W_{Λ} cares only about the magnitude of the given k -mode, we can rewrite Eq. (3.45) as

$$\begin{aligned} \left\langle \left(\delta_{\Lambda}^{(1)} \right)^2 \left(W_{\Lambda} \left(\delta_{\Lambda_0}^{(1)} \right)^2 \right) \right\rangle(\mathbf{k}) &= 2 \int_{\mathbf{k}'} W_{\Lambda}(|\mathbf{k} - \mathbf{k}'|) W_{\Lambda}(k') P_L(|\mathbf{k} - \mathbf{k}'|) P_L(k') \\ &\quad + \int_{\mathbf{k}_1}^{k=0} W_{\Lambda}(k_1) P_L(k_1) \int_{\mathbf{k}_2} W_{\Lambda_0}(k_2) P_L(k_2), \end{aligned} \quad (3.46)$$

indicating that the second line only contributes to the $\mathbf{k} = 0$ mode, which is not included in the likelihood evaluation, so this line can be dropped from further calculation. Therefore, the only relevant piece is the loop integral on the first line, which in fact matches the correlator on the right-hand side of Eq. (3.44). This shows that $b_{\delta^2, \Lambda_0} = b_{\delta^2, \Lambda}$, and hence no running of b_{δ^2} is expected for the forward model represented by Eq. (3.5). This behavior is confirmed within our inference chains in Fig. 3.6. Note that this result is specific to this simple forward model, and does not apply to the n LPT forward models.

Forward model, Λ [$h \text{ Mpc}^{-1}$]	Dataset	$(\hat{R}_C - 1) \times 10^3$	$(\hat{R}_r - 1, \mathcal{T}_\epsilon - 1) \times 10^4$	\hat{E}_s
LINEAR - Eq. (3.3), $\Lambda = 0.05$	$\mathcal{D}_1^{\text{LINEAR}}$	5.3	(129.8, 7.97)	113
LINEAR - Eq. (3.3), $\Lambda = 0.07$	$\mathcal{D}_1^{\text{LINEAR}}$	29.6	(15.65, 7.97)	320
LINEAR - Eq. (3.3), $\Lambda = 0.1$	$\mathcal{D}_1^{\text{LINEAR}}$	1.79	(21.79, 7.97)	230
LINEAR - Eq. (3.3), $\Lambda = 0.14$	$\mathcal{D}_1^{\text{LINEAR}}$	2.31	(48.79, 7.97)	102
LINEAR - Eq. (3.5), $\Lambda = 0.05$	$\mathcal{D}_{2,a}^{\text{LINEAR}}$	10.2	(4.29, 7.38)	1138
LINEAR - Eq. (3.5), $\Lambda = 0.05$	$\mathcal{D}_{2,b}^{\text{LINEAR}}$	15.2	(5.26, 7.38)	933
LINEAR - Eq. (3.5), $\Lambda = 0.08$	$\mathcal{D}_{2,a}^{\text{LINEAR}}$	2.66	(4.64, 7.38)	1051
LINEAR - Eq. (3.5), $\Lambda = 0.08$	$\mathcal{D}_{2,b}^{\text{LINEAR}}$	4.14	(4.54, 7.38)	1075
LINEAR - Eq. (3.5), $\Lambda = 0.1$	$\mathcal{D}_{2,a}^{\text{LINEAR}}$	1.62	(7.69, 7.38)	635
LINEAR - Eq. (3.5), $\Lambda = 0.1$	$\mathcal{D}_{2,b}^{\text{LINEAR}}$	4.26	(7.86, 7.38)	622
LINEAR - Eq. (3.5), $\Lambda = 0.12$	$\mathcal{D}_{2,a}^{\text{LINEAR}}$	8.71	(30.18, 7.38)	330
LINEAR - Eq. (3.5), $\Lambda = 0.12$	$\mathcal{D}_{2,b}^{\text{LINEAR}}$	11.21	(35.96, 7.38)	277
LINEAR - Eq. (3.5), $\Lambda = 0.13$	$\mathcal{D}_{2,a}^{\text{LINEAR}}$	13.84	(42.91, 7.38)	232
LINEAR - Eq. (3.5), $\Lambda = 0.13$	$\mathcal{D}_{2,b}^{\text{LINEAR}}$	10.74	(34.56, 7.38)	288
LINEAR - Eq. (3.5), $\Lambda = 0.14$	$\mathcal{D}_{2,a}^{\text{LINEAR}}$	24.9	(33.69, 7.38)	147
LINEAR - Eq. (3.5), $\Lambda = 0.14$	$\mathcal{D}_{2,b}^{\text{LINEAR}}$	14.8	(25.31, 7.38)	195
1LPT - Eq. (3.11), $\Lambda = 0.1$	$\mathcal{D}_{1,a}^{2\text{LPT}}$	3.25	(38.38, 6.89)	256
1LPT - Eq. (3.11), $\Lambda = 0.1$	$\mathcal{D}_{1,b}^{2\text{LPT}}$	5.37	(73.68, 6.89)	135
1LPT - Eq. (3.11), $\Lambda = 0.14$	$\mathcal{D}_{1,a}^{2\text{LPT}}$	41.8	(28.39, 6.89)	171
1LPT - Eq. (3.11), $\Lambda = 0.14$	$\mathcal{D}_{1,b}^{2\text{LPT}}$	23.2	(42.95, 6.89)	114
2LPT - Eq. (3.11), $\Lambda = 0.1$	$\mathcal{D}_{1,a}^{2\text{LPT}}$	2.46	(36.32, 6.89)	273
2LPT - Eq. (3.11), $\Lambda = 0.1$	$\mathcal{D}_{1,b}^{2\text{LPT}}$	23.5	(40.71, 6.89)	243
2LPT - Eq. (3.11), $\Lambda = 0.14$	$\mathcal{D}_{1,a}^{2\text{LPT}}$	17.5	(31.66, 6.89)	311
2LPT - Eq. (3.11), $\Lambda = 0.14$	$\mathcal{D}_{1,b}^{2\text{LPT}}$	95.1	(25.56, 6.89)	385
2LPT - Eq. (3.11), $\Lambda = 0.1$	$\mathcal{D}_{2,a}^{2\text{LPT}}$	2.78	(53.73, 6.89)	184
2LPT - Eq. (3.11), $\Lambda = 0.1$	$\mathcal{D}_{2,b}^{2\text{LPT}}$	0.17	(63.17, 6.89)	157
2LPT - Eq. (3.11), $\Lambda = 0.14$	$\mathcal{D}_{2,a}^{2\text{LPT}}$	94.3	(51.18, 6.89)	125
2LPT - Eq. (3.11), $\Lambda = 0.14$	$\mathcal{D}_{2,b}^{2\text{LPT}}$	99.6	(47.42, 6.89)	103

Table 3.2: Gelman-Rubin test statistics, both \hat{R}_C (see Eq. (3.51)) and \hat{R}_r (see Eq. (3.52)) for our MCMC chains using the unmarginalized likelihood presented in this work. We also show the idealized convergence threshold value, \mathcal{T}_ϵ , corresponding to having the same number of chains, but instead requiring that 95% of samples lie within $< 10\%$ of the posterior volume around the posterior mean which is reported by the `target.psrff` method of `stableGR` package (see [125]). The last column indicates the estimated effective sample size, \hat{E}_s , calculated using the `n.eff` method of `stableGR`. The results in each line are obtained from at least 2 chains. For calculating the \hat{R}_C value, we considered b_δ parameter in case of forward models from Section 3.5.1, and α parameter for the chains from Section 3.5.2 (see text for more details about this choice).

Forward model, $\Lambda [h \text{ Mpc}^{-1}]$	Dataset	$(\hat{R}_C - 1) \times 10^3$	$(\hat{R}_r - 1, \mathcal{T}_\epsilon - 1) \times 10^4$	\hat{E}_s
LINEAR - Eq. (3.5), $\Lambda = 0.08$	$\mathcal{D}_{2,a}^{\text{LINEAR-i.}}$	0.21	(3.24, 7.38)	1516
LINEAR - Eq. (3.5), $\Lambda = 0.08$	$\mathcal{D}_{2,b}^{\text{LINEAR-i.}}$	0.22	(3.49, 7.38)	1408
LINEAR - Eq. (3.5), $\Lambda = 0.1$	$\mathcal{D}_{2,a}^{\text{LINEAR-i.}}$	31.4	(10.58, 7.38)	469
LINEAR - Eq. (3.5), $\Lambda = 0.1$	$\mathcal{D}_{2,b}^{\text{LINEAR-i.}}$	3.54	(9.49, 7.38)	523
LINEAR - Eq. (3.5), $\Lambda = 0.12$	$\mathcal{D}_{2,a}^{\text{LINEAR-i.}}$	9.45	(22.26, 7.38)	224
LINEAR - Eq. (3.5), $\Lambda = 0.12$	$\mathcal{D}_{2,b}^{\text{LINEAR-i.}}$	0.89	(16.81, 7.38)	296
LINEAR - Eq. (3.5), $\Lambda = 0.13$	$\mathcal{D}_{2,a}^{\text{LINEAR-i.}}$	31.81	(37.81, 7.38)	132
LINEAR - Eq. (3.5), $\Lambda = 0.13$	$\mathcal{D}_{2,b}^{\text{LINEAR-i.}}$	36.64	(41.13, 7.38)	121
LINEAR - Eq. (3.5), $\Lambda = 0.14$	$\mathcal{D}_{2,a}^{\text{LINEAR-i.}}$	9.95	(24.46, 7.38)	205
LINEAR - Eq. (3.5), $\Lambda = 0.14$	$\mathcal{D}_{2,b}^{\text{LINEAR-i.}}$	9.64	(23.57, 7.38)	211

Table 3.3: Supplement table for Tab. 3.2 containing Gelman-Rubin test statistics for chains obtained from applying the forward model from Eq. (3.5) to the $\mathcal{D}_2^{\text{LINEAR-informative}}$ datasets appearing in Fig. 3.6. We abbreviate $\mathcal{D}_2^{\text{LINEAR-informative}} \rightarrow \mathcal{D}_2^{\text{LINEAR-i.}}$ to save space.

Forward model, $\Lambda [h \text{ Mpc}^{-1}]$	Dataset	$(\hat{R}_C - 1) \times 10^3$	$(\hat{R}_r - 1, \mathcal{T}_\epsilon - 1) \times 10^4$	\hat{E}_s
1LPT - Eq. (3.11), $\Lambda = 0.1$	$\mathcal{D}_{1,a}^{2\text{LPT}}$	6.75	(116.51, 7.97)	128
1LPT - Eq. (3.11), $\Lambda = 0.1$	$\mathcal{D}_{1,b}^{2\text{LPT}}$	19.5	(137.9, 7.97)	108
1LPT - Eq. (3.11), $\Lambda = 0.14$	$\mathcal{D}_{1,a}^{2\text{LPT}}$	22.5	(96.3, 7.97)	155
1LPT - Eq. (3.11), $\Lambda = 0.14$	$\mathcal{D}_{1,b}^{2\text{LPT}}$	44.9	(102.9, 7.97)	145
2LPT - Eq. (3.11), $\Lambda = 0.1$	$\mathcal{D}_{1,a}^{2\text{LPT}}$	4.78	(60.61, 7.97)	166
2LPT - Eq. (3.11), $\Lambda = 0.1$	$\mathcal{D}_{1,b}^{2\text{LPT}}$	4.54	(63.89, 7.97)	156
2LPT - Eq. (3.11), $\Lambda = 0.14$	$\mathcal{D}_{1,a}^{2\text{LPT}}$	99.5	(80.76, 7.97)	185
2LPT - Eq. (3.11), $\Lambda = 0.14$	$\mathcal{D}_{1,b}^{2\text{LPT}}$	28.3	(85.84, 7.97)	174
2LPT - Eq. (3.11), $\Lambda = 0.1$	$\mathcal{D}_{2,a}^{2\text{LPT}}$	98.4	(108.9, 7.96)	137
2LPT - Eq. (3.11), $\Lambda = 0.1$	$\mathcal{D}_{2,b}^{2\text{LPT}}$	94.2	(115.6, 7.97)	129
2LPT - Eq. (3.11), $\Lambda = 0.14$	$\mathcal{D}_{2,a}^{2\text{LPT}}$	6.47	(113.86, 7.97)	131
2LPT - Eq. (3.11), $\Lambda = 0.14$	$\mathcal{D}_{2,b}^{2\text{LPT}}$	78.8	(51.95, 7.97)	192

Table 3.4: Same as Tab. 3.2, but for chains using the marginalized likelihood from Eq. (3.16). Note that these chains were not run as long as the chains from Tab. 3.2 and hence have a smaller number of effective samples overall.

3.10 Appendix D: Convergence and sample correlation

MCMC samples are not entirely independent. In practice, this correlation between neighboring samples introduces further uncertainty in any estimate based on averaging over those samples, such as the posterior mean, variance and all higher moments. The correlation is often measured by the integrated autocorrelation time, while the resulting uncertainty in the posterior is quantified by the effective sample size. We refer readers to [124] for more details.

We use the following definition of normalized autocorrelation function, $\rho(t)$

$$\begin{aligned}\rho(t) &\equiv \frac{\mathcal{A}(t)}{\mathcal{A}(0)}, \\ \mathcal{A}(t) &= \langle \gamma_s \gamma_{s+t} \rangle_s - \langle \gamma_s \rangle_s^2,\end{aligned}\tag{3.47}$$

where $\{\gamma_s\}_{s=1\dots N}$ is the set of chain samples, and the brackets indicate the average over samples, i.e. $\langle \gamma_s \rangle_s \equiv \bar{\gamma}$, while t indicates the sample separation. Eq. (3.47) highlights the significance of having a sufficient number of MCMC samples, i.e. running sufficiently long MCMC chains, since $\bar{\gamma}$ and $\mathcal{A}(0)$ are noisy estimates of the true mean and variance, whose noise propagates nonlinearly into $\rho(t)$. For the autocorrelation function, $\mathcal{A}(t)$, we use the estimator presented in [129] which shows better asymptotic behavior than the one given in Eq. (3.47). The normalized autocorrelation function $\rho(t)$ is exactly what shown in Fig. 3.9. We then use the autocorrelation function to estimate the correlation length of the chain as

$$\hat{\tau}(T) = \sum_{t=-T}^T \rho(t) = 1 + 2 \sum_{t=1}^T \rho(t),\tag{3.48}$$

with T representing the maximal separation between the samples considered. This estimator has a vanishing variance in the limit of large chain lengths, i.e. number of samples. We have adopted the approach of [130] for choosing T . In short, T is chosen such that it corresponds to the smallest integer satisfying $T \geq C\hat{\tau}(T)$ with a constant C chosen such that the variance of the estimator is minimized, at the cost of introducing a negative bias in the estimate of $\hat{\tau}$. This is typically achieved for $C \in [5, 10]$. We report the $\hat{\tau}$ estimates, as well as the used window T in Fig. 3.9 as well.

We now describe the two tests of convergence we perform for all chains analyzed in this work, namely the classical and revised Gelman-Rubin (G-R) diagnostic. The revised G-R statistics [125]) makes a clear connection to the chain effective sample size (see Eq. (12) in [125]). We exploit this connection to link the (revised) G-R value and our target effective sample size.

For the classical G-R statistics [123, 124, 131], we adopt the following procedure.

First we calculate the inter- and intra-chain variances

$$\begin{aligned} B &= \frac{N}{M-1} \sum_{j=1}^M (\bar{\gamma}_{.j} - \bar{\gamma}_{..})^2, \\ W &= \frac{1}{M} \sum_{j=1}^M s_j^2, \end{aligned} \quad (3.49)$$

with

$$\bar{\gamma}_{.j} = \frac{1}{N} \sum_{i=1}^N \gamma_{ij}, \quad \bar{\gamma}_{..} = \frac{1}{M} \sum_{j=1}^M \bar{\gamma}_{.j}, \quad s_j^2 = \frac{1}{N-1} \sum_{i=1}^N (\gamma_{ij} - \bar{\gamma}_{.j})^2,$$

and N, M being the chain length and number of chains considered respectively. From the above expression one can see that B represents an estimate of the variance between the chains while W is the mean of the variance within individual chains. These two quantities then can be combined into an estimate of the true underlying target distribution variance

$$\hat{\sigma}^2 = \frac{N-1}{N} W + \frac{1}{N} B. \quad (3.50)$$

The authors of [123, 131] argue that for a properly dispersed set of chains, the $\hat{\sigma}^2$ estimate is typically over-estimating the underlying variance, while the mean of the within-the-chain variances, W , under-estimates it. Hence, they propose the following quantity as a measure of chain convergence, which is known as the classical G-R test statistics

$$\hat{R}_C = \sqrt{\frac{\hat{\sigma}^2}{W}}. \quad (3.51)$$

Specifically, we apply this univariate G-R test to the α parameter (for chains from Section 3.4.2) and b_δ parameter (for chains from Section 3.5.1), since these parameters typically have the longest correlation lengths and largest \hat{R}_C values. The resulting values are reported in Tab. 3.2, 3.3 and 3.4 as \hat{R}_C .

The revised G-R test statistics can be estimated from the following expression

$$\hat{R}_r \approx \sqrt{1 + \frac{M}{\hat{E}_s}} \leq \mathcal{T}_\epsilon. \quad (3.52)$$

The above Eq. (3.52) provides a clear connection between the number of chains M , effective sample size \hat{E}_s and the convergence threshold \mathcal{T}_ϵ . It is possible to determine the convergence threshold a-priori and hence the corresponding effective size necessary for reaching it. Specifically, in this work, we set a target of $\hat{E}_s \geq 100$ for all MCMC

chains. In practice, we use the `n.eff` multivariate method of the `stable.GR` package⁵ to estimate \hat{E}_s . The `stable.GR` package is provided by the authors of [125]. For more details on how \hat{E}_s is calculated, we refer readers to Section 5 of [125]. This number is reported in the last column of Tab. 3.2, 3.3 and 3.4. We note that the convergence threshold value \mathcal{T}_ϵ we report represents an ideal case, which corresponds to having the same number of chains as we do, but with 95% of samples lying within $< 10\%$ of the posterior volume around the posterior mean. Such chains will then have $\hat{R}_r \sim \mathcal{T}_\epsilon$.

3.11 Appendix E: 1LPT and 2LPT 2nd order bias coefficients

As described in Section 3.2, the 1LPT and 2LPT forward models differ only in the order of the LPT displacement field. This difference has an impact on the inference as shown in Fig. 3.8. The goal of this section is to understand whether the observed discrepancy between the two chains is expected. In order to derive this, we first go back to the general setup of both of these models.

First, the same order of Lagrangian bias expansion is employed in both, allowing for the following set of Lagrangian bias operators

$$O_L \in \left\{ 1, (\text{tr } M^{(1)})^2, \text{tr } (M^{(1)} M^{(1)}) \right\}, \quad (3.53)$$

and the corresponding bias coefficients. Recall that we displace a unit field to obtain the Eulerian matter density. The transformation of bias coefficients derived in the following only involves operators at leading order in derivatives, therefore we do not need to consider $\nabla^2 \delta_{n\text{-LPT}}$ in the following. The relation between Eulerian and Lagrangian frames is given by [31]

$$\begin{aligned} 1 + \delta_{\text{det},\Lambda}^{n\text{LPT}}(\mathbf{x}) &= (\mathcal{J}^{(n)})^{-1} (1 + \delta_{\text{det},\Lambda}^{n\text{LPT}}(\mathbf{q})), \\ \mathcal{J}^{(n)} &= \det(\mathbb{1} + \partial_{\mathbf{q}} \boldsymbol{\psi}^{(n)}), \end{aligned} \quad (3.54)$$

with \mathcal{J} being the Jacobian of the transformation from Lagrangian to Eulerian coordinates, and n denoting the order up to which the forward model is to be evaluated. Note that in our field-level forward model, \mathcal{J}^{-1} is computed non-perturbatively, by displacing and depositing pseudo particles within the simulated box. Here, we instead expand \mathcal{J}^{-1} perturbatively up to second order, to obtain the mapping of bias operators between the different LPT orders. Specifically, for 1LPT and 2LPT the corresponding inverse

⁵<https://github.com/knudson1/stableGR>

Jacobians are given by

$$\begin{aligned} (\mathcal{J}^{(1)})^{-1} &= 1 - \partial_{q_i} \psi_i^{(1)} + \frac{1}{2} \left[\left(\partial_{q_i} \psi_i^{(1)} \right)^2 + \partial_{q_i} \psi_j^{(1)} \partial_{q_j} \psi_i^{(1)} \right] + o((\psi^{(1)})^3), \\ (\mathcal{J}^{(2)})^{-1} &= 1 - \partial_{q_i} \psi_i^{(1)} - \partial_{q_j} \psi_j^{(2)} + \frac{1}{2} \left[\left(\partial_{q_i} \psi_i^{(1)} \right)^2 + \partial_{q_i} \psi_j^{(1)} \partial_{q_j} \psi_i^{(1)} \right] + o((\psi^{(1)})^3), \end{aligned} \quad (3.55)$$

where we kept only second order terms. We then use the following bias expansion

$$\delta_{\text{det},\Lambda}(\mathbf{q}) = b_\delta^L + b_{\sigma^2}^L (\text{tr } M^{(1)})^2 + b_{\text{tr}[M^{(1)}M^{(1)}]}^L \text{tr} (M^{(1)}M^{(1)}), \quad (3.56)$$

where we note that b_δ^L is the coefficient of the uniform field $O_L = 1$. Plugging the results of Eqs. (3.55)–(3.56) directly into the first line of Eq. (3.54) we obtain

$$\begin{aligned} \delta_{\text{det},\Lambda}^{1\text{LPT}}(\mathbf{x}) &= (b_\delta^L + 1)(1 - \text{tr } M^{(1)}) + \left(b_{\sigma^2}^L + \frac{1}{2}(b_\delta^L + 1) \right) (\text{tr } M^{(1)})^2 \\ &\quad + \left(b_{\text{tr}[M^{(1)}M^{(1)}]}^L + \frac{1}{2}(b_\delta^L + 1) \right) \text{tr} (M^{(1)}M^{(1)}), \\ \delta_{\text{det},\Lambda}^{2\text{LPT}}(\mathbf{x}) &= (b_\delta^L + 1)(1 - \text{tr } M^{(1)}) + \left(b_{\sigma^2}^L + \frac{1}{2}(b_\delta^L + 1) - \frac{3}{14} \right) (\text{tr } M^{(1)})^2 \\ &\quad + \left(b_{\text{tr}[M^{(1)}M^{(1)}]}^L + \frac{1}{2}(b_\delta^L + 1) + \frac{3}{14} \right) \text{tr} (M^{(1)}M^{(1)}), \end{aligned} \quad (3.57)$$

again keeping only second order terms. In Eq. (3.57) we also used the solution for the second order displacement field from the equations of motion assuming an Einstein-de Sitter universe (see [117], as well as Section 2.7 in [55], Section 2.5.2 in [31])

$$\text{tr } M^{(2)} = -\frac{3}{14} \left((\text{tr } M^{(1)})^2 - \text{tr} (M^{(1)}M^{(1)}) \right).$$

This then produces the following relationship between the 1LPT and 2LPT forward model bias coefficients

$$\begin{aligned} b_\delta^{2\text{LPT}} &= b_\delta^{1\text{LPT}} = b_\delta^L + 1, \\ b_{\sigma^2}^{1\text{LPT}} &= b_{\sigma^2}^{2\text{LPT}} + \frac{3}{14} \\ b_{\text{tr}[M^{(1)}M^{(1)}]}^{1\text{LPT}} &= b_{\text{tr}[M^{(1)}M^{(1)}]}^{2\text{LPT}} - \frac{3}{14}. \end{aligned} \quad (3.58)$$

We can now compare with the results from the 1LPT and 2LPT inferences on the $\mathcal{D}_1^{2\text{LPT}}$ datasets, shown in Fig. 3.8. Taking the mean of the obtained 2LPT posterior for $b_{\sigma^2}^{2\text{LPT}} \approx 0.0236$ and $b_{\text{tr}[M^{(1)}M^{(1)}]}^{2\text{LPT}} \approx -0.099$, leads to $b_{\sigma^2}^{1\text{LPT}} \approx 0.238$ and $b_{\text{tr}[M^{(1)}M^{(1)}]}^{1\text{LPT}} \approx -0.31$, which are indicated with dotted lines in Fig. 3.8, which agree within 68 – 95% confidence level with the obtained 1LPT posterior.

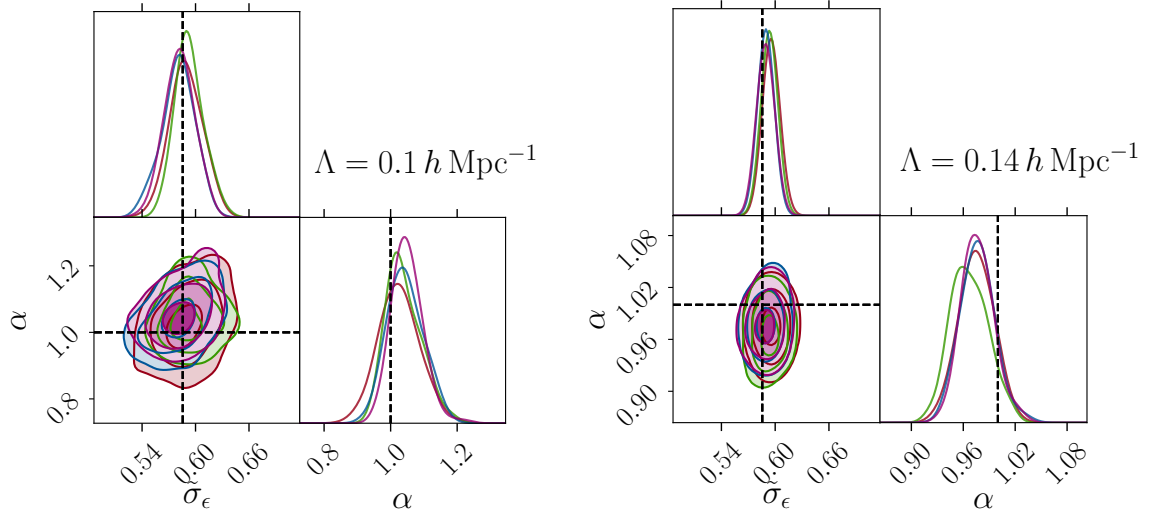


Figure 3.12: Parameter posteriors obtained using the marginalized (Eq. (3.16)) and unmarginalized likelihoods (Eq. (3.15)) in the $\alpha - \sigma_\epsilon$ plane, both inferred from the $\mathcal{D}_{1,b}^{2\text{LPT}}$ dataset. We show the results at different cutoffs, on the left, $\Lambda = 0.1 h \text{ Mpc}^{-1}$ and on the right $\Lambda = 0.14 h \text{ Mpc}^{-1}$. Forward models considered are 1LPT with unmarginalized likelihood (green), 1LPT with marginalized likelihood (red), 2LPT with unmarginalized likelihood (purple) and 2LPT with marginalized likelihood (blue). There is also an excellent agreement among the two likelihoods, while the marginalized likelihood yields a smaller correlation length for the α parameter (see Fig. 3.9).

3.12 Appendix F: Marginalized and unmarginalized likelihood

In this section, we compare the posteriors obtained with the marginalized likelihood (Eq. (3.16)), using the priors given in Eq. (3.19), and unmarginalized (Eq. (3.15)) likelihood, using priors given in Eq. (3.18), both inferred from the $\mathcal{D}_{1,b}^{2\text{LPT}}$ dataset. Results are the same for the inference done on $\mathcal{D}_{1,a}^{2\text{LPT}}$ dataset, but for the sake of clarity, we just show the inference on $\mathcal{D}_{1,b}^{2\text{LPT}}$.

Results are shown in Fig. 3.12. We also note that inferences on $\mathcal{D}_1^{2\text{LPT}}$ and $\mathcal{D}_2^{2\text{LPT}}$ datasets from Section 3.5.2 show similar trends. That is, the posteriors obtained with unmarginalized and marginalized likelihood agree.

As can be seen in Fig. 3.12, the posterior contours are entirely consistent with each other for both inferences with 1LPT and 2LPT) forward models, and across the different cutoffs ($\Lambda = 0.1 h \text{ Mpc}^{-1}$ and $\Lambda = 0.14 h \text{ Mpc}^{-1}$). However, recall that the marginalized likelihood leads to a better efficiency in terms of sampling the α parameter, as depicted in Fig. 3.9 (top panel).

CHAPTER 4

CONSTRAINING WIMP DARK MATTER USING THE LOCAL UNIVERSE

The following is a summary of results presented from the following papers [132, 133] with Deaglan Bartlett, Harry Desmond, Jens Jasche and Guilhem Lavaux as co-authors. First paper [132] is with me as second author, and the second paper [133] is submitted to Physical Review D with me as first author.

Abstract

Decaying or annihilating dark matter particles could be detected through gamma-ray emission from the species they decay or annihilate into. This is usually done by modelling the flux from specific dark matter-rich objects such as the Milky Way halo, Local Group dwarfs, and nearby groups. However, these objects are expected to have significant emission from baryonic processes as well, and the analyses discard gamma-ray data over most of the sky. Here we construct full-sky templates for gamma-ray flux from the large-scale structure within ~ 200 Mpc by means of a suite of constrained N -body simulations (CSiBORG) produced using the BORG algorithm. Marginalising over uncertainties in this reconstruction, small-scale structure, and parameters describing astrophysical contributions to the observed gamma-ray sky, we compare to observations from the *Fermi* Large Area Telescope to constrain dark matter annihilation cross sections and decay rates through a Markov Chain Monte Carlo analysis. We rule out the thermal relic cross section, $\sigma v = a_\ell (v/c)^{2\ell}$, for s -wave ($\ell = 0$) annihilation for all $m_\chi \lesssim 7 \text{ GeV}/c^2$ at 95% confidence if the annihilation produces gluons or quarks less massive than the bottom quark. We infer a contribution to the gamma-ray sky with the same spatial distribution as dark matter decay at 3.3σ . Although this could be due to dark matter decay via

these channels with a decay rate $\Gamma \approx 6 \times 10^{-28} \text{ s}^{-1}$, we find that a power-law spectrum of index $p = -2.75_{-0.46}^{+0.71}$, likely of baryonic origin, is preferred by the data. Furthermore, we also derive constraints for annihilation channels with velocity dependent cross sections. We are able to obtain constraints on a_ℓ , for both p -wave ($\ell = 1$) and d -wave ($\ell = 2$) channels, which are two and seven orders of magnitude tighter than those obtained from dwarf spheroidals respectively. We find no evidence for p -wave or d -wave annihilation from dark matter particles with masses in the range $m_\chi = 2 - 500 \text{ GeV}/c^2$. In particular, for annihilations producing bottom quarks with $m_\chi = 10 \text{ GeV}/c^2$, we find $a_1 < 2.4 \times 10^{-21} \text{ cm}^3\text{s}^{-1}$ and $a_2 < 3.0 \times 10^{-18} \text{ cm}^3\text{s}^{-1}$ at 95% confidence, although we note that these constraints are much weaker than the corresponding thermal relic cross sections.

4.1 Introduction

An unambiguous detection of non-gravitational interactions of dark matter continues to evade us. Astrophysical objects are ideal targets in the search for novel dark matter physics, since low interaction rates can be compensated by the large quantities of dark matter, leading to potentially detectable signals. For over a decade, the origin of the observed excess of gamma rays at the centre of our galaxy – known as the Galactic Centre Excess (GCE) [134–136] – has been debated, with explanations such as unresolved point sources or dark matter annihilation motivated by various extensions of the Standard Model (SM) being proposed, e.g. [137–163]. Moreover, experiments suggest an anomalous muon magnetic moment, as measured by the “Muon $g - 2$ ” experiment [164], and a mass for the W boson that is higher than expected [165], which further motivates probing physics beyond SM and thus the search for dark matter (DM) candidates. Detection of such particles could therefore solve some of the fundamental questions of particle physics and cosmology. Favourable candidates are weakly interacting massive particles (WIMPs) [57, 166, 167] (see also Section 2.2), including, but not limited to, the lightest supersymmetric particle in supersymmetric theories. Given the conflicting explanations of the GCE, and motivated by the search of beyond SM physics, one should determine if such models are consistent with detections from other sources or across the full sky.

So far, the majority of work on GCE excess has focused on s -wave annihilation, where the product of the self-annihilation cross section, σ , and relative velocity, v , is independent of v . Such models can be strongly constrained by considering the gamma ray flux from nearby dwarf galaxies [168–170], which are able to rule out the thermal relic cross section for dark matter particle masses relevant to explain the GCE. However, not all dark matter annihilation models require σv to be independent of velocity. For example, for fermionic dark matter annihilation to spin-0 particles, even parity final states cannot have an s -wave contribution in parity conserving theories [171, 172]. p -

wave annihilation – where σv is proportional to the square of the relative velocity – dominates when Standard Model fermion-antifermion pairs are produced from Majorana fermions in models with minimal flavour violation due to chirality suppression of s -wave annihilation [173]. If dark matter is instead a real scalar, then d -wave annihilation – where the fourth power of the relative velocity is relevant – is dominant [174, 175].

In our work [132, 133], instead of focusing on a few nearby or massive objects, the aim is to search for the signature of DM decay and annihilation across the full sky by forward-modelling the observed gamma-ray sky, as first suggested in [176]. As proposed in [177], anisotropies in the cosmic gamma-ray background could be a signature of DM annihilation or decay. This has previously been studied through the two-point correlation function, for example [178], which is calibrated with *unconstrained* N -body simulations (e.g. [179] use the Millennium-II and Aquarius simulations [180–182]). Instead, we utilise the CSiBORG suite of *constrained* N -body simulations [183–186]. The initial conditions (ICs) for these simulations are chosen to produce final three-dimensional DM density fields that are consistent with the observed positions of galaxies in the 2M++ galaxy catalogue. The ICs are inferred using the BORG (see also Section 2.6.1) algorithm [19–22, 187], a fully Bayesian forward model. We use the particle positions in the simulations to produce maps of the expected gamma-ray flux from halos down to $4.38 \times 10^{11} M_{\odot}$ in mass (although we also model smaller substructures), as well as from DM not identified as belonging to halos.

For our fiducial analyses, the halos are assumed to have Navarro-Frenk-White (NFW) profiles [188]. For our s -wave analysis [132], we explicitly model unresolved substructure in a probabilistic manner, since the signal from DM annihilation is sensitive to the peaks in the density field. On the other hand, for our p - and d -wave fiducial analyses [133], we don’t marginalize over the substructure contribution given its negligible effect for these two channels (see for example [189], although note that this conclusion does not hold for the case of Sommerfeld enhanced annihilation [190, 191]). In contrast to the s -wave case, for the velocity dependent cross sections, the full-sky analysis has a particular advantage over local dwarf spheroidal sources, since one would expect a large signal from extra-galactic halos, where the velocity dispersion should be larger. This is in line with what has recently been suggested to use extra-galactic halos to search for velocity-dependent dark matter annihilation signals in gamma ray data [192].

In our s -wave analysis, we are able to rule out the thermal relic cross section at 95% confidence for annihilations that produce gluons or quarks less massive than the bottom quark if DM has a mass $m_{\chi} \lesssim 7 \text{ GeV}/c^2$. We find a contribution to the gamma-ray sky with the same spatial distribution as expected from DM decay (flux proportional to DM density) at 3.3σ confidence, with a decay rate $\Gamma \approx 6 \times 10^{-28} \text{ s}^{-1}$ for these channels. However, a power-law spectrum with an index $p = -2.75_{-0.46}^{+0.71}$ provides a better fit to the data, suggesting a non-DM origin. For s -wave, we are not able to rule out the thermal relic annihilation cross section at any mass for production of top or bottom quarks; we

obtain upper bounds that are half the size (i.e. tighter constraints) if we marginalise over the contribution proportional to the DM density, but we do not include this contribution in the fiducial analysis. Our constraints on DM decay to leptons are approximately an order of magnitude less stringent than decay to quarks.

Regarding the p - and d -wave cases, we find no evidence for either type of annihilation from dark matter particles with masses in the range $m_\chi = 2 - 500 \text{ GeV}/c^2$, for any channel. The bounds we obtain, however fail to exclude the thermal relic cross section for these velocity-dependent annihilation channels. Nonetheless, the derived constraints are among the tightest to date.

4.1.1 Theoretical background

In this section we elaborate on the calculation of the flux templates for our annihilation and decay models. First we describe how we treat the s -wave annihilation and decay case and then immediately after the velocity dependent annihilation cases for p - and d -wave.

Velocity independent case

We start by assuming that DM is made of a single particle, χ , of mass m_χ , whose antiparticle is itself (e.g., Majorana fermions). This particle is assumed to be able to both decay

$$\chi \rightarrow A\bar{A}, \quad (4.1)$$

and annihilate

$$\chi\chi \rightarrow A\bar{A}, \quad (4.2)$$

to a standard model particle, A , and its antiparticle, \bar{A} . The annihilation of the produced particles would lead to gamma-ray emission at some energy E_γ , which one could detect. If these processes occur via channel i with branching ratio Br_i , then the photon flux for annihilation per unit density squared at redshift z is [193]

$$\frac{d\Phi_\gamma^{\text{ann}, \ell=0}}{dE_\gamma} = \frac{\langle\sigma v\rangle}{8\pi m_\chi^2} \sum_i \text{Br}_i \left. \frac{dN_i^{\text{ann}}}{dE'_\gamma} \right|_{E'_\gamma=E_\gamma(1+z)}, \quad (4.3)$$

and for decay per unit density

$$\frac{d\Phi_\gamma^{\text{dec}}}{dE_\gamma} = \frac{\Gamma}{4\pi m_\chi} \sum_i \text{Br}_i \left. \frac{dN_i^{\text{dec}}}{dE'_\gamma} \right|_{E'_\gamma=E_\gamma(1+z)}, \quad (4.4)$$

where $\langle\sigma v\rangle$ is the thermally averaged cross section, $\tau = 1/\Gamma$ is the lifetime of the particle, and $dN_i^{(\cdot)}/dE_\gamma$ is the photon energy distribution for channel i . Note that the photon distribution function in both cases, $dN_i^{(\cdot)}/dE'_\gamma$, should be evaluated at the redshift of the

halo. Since we will be using low-redshift halos in this work ($z \lesssim 0.05$) we will neglect this redshift effect and thus assume that $E'_\gamma \approx E_\gamma$. Introducing redshift dependence would require inserting a redshift dependent optical depth into Eq. (4.10), which, in our case, will only marginally affect the results (see Section 4.5.2). For simplicity, we assume that the decays occur via one channel, so the branching ratio, Br_i is either 0 or 1. The superscript $\ell = 0$ reminds the reader that this equation holds only for the case of s -wave annihilation, which assumes velocity independent cross section. The generalization to a velocity dependent annihilation-channel such as the p - and d -wave, will be presently shortly.

Now, if χ is not its own antiparticle (e.g., Dirac fermions), the annihilation flux is half of this value, provided there is no matter-antimatter asymmetry. Since we do not *a priori* know which branching ratios to use, in this work we assume that the annihilation or decay occurs via a single channel; however, our analysis can be trivially extended to multiple channels.

The above results apply at unit density, we must now take into account the integrated DM density along the line of sight. By introducing the J factor

$$\frac{dJ^{\ell=0}}{d\Omega} = \int \rho_{\text{DM}}^2(\mu, \Omega) d\mu, \quad (4.5)$$

where μ is a radial line of sight coordinate, and the D factor

$$\frac{dD}{d\Omega} = \int \rho_{\text{DM}}(\mu, \Omega) d\mu, \quad (4.6)$$

we arrive at the total photon fluxes per unit solid angle

$$\frac{d^2\Phi^{\text{ann}}}{dE_\gamma d\Omega} = \frac{d\Phi_\gamma^{\text{ann}}}{dE_\gamma} \frac{dJ}{d\Omega}, \quad (4.7)$$

$$\frac{d^2\Phi^{\text{dec}}}{dE_\gamma d\Omega} = \frac{d\Phi_\gamma^{\text{dec}}}{dE_\gamma} \frac{dD}{d\Omega}, \quad (4.8)$$

where we note that we have assumed that the cosmological redshift variation across the source is negligible, so we can factor out Eqs. (4.3) and (4.4) from the line of sight integral.

Velocity dependent case

We also consider dark matter annihilation with a cross section σ which depends on the relative velocity, v , between annihilating dark matter particles. We can therefore generalize and consider the following relation

$$\sigma v = a_\ell S_\ell \left(\frac{v}{c} \right), \quad (4.9)$$

where a_ℓ is a constant and S_ℓ can be an arbitrary function of v/c . At observed energy E_γ , the differential photon flux, $d\Phi_\gamma^{\text{ann},\ell}/dE_\gamma$, is

$$\frac{d\Phi_\gamma^{\text{ann},\ell}}{dE_\gamma} = \frac{a_\ell}{8\pi m_\chi^2} J^\ell \sum_i \text{Br}_i \frac{dN_i^{\text{ann}}}{dE_\gamma} \quad (4.10)$$

where we sum over annihilation channels with branching ratios Br_i . This equation is valid if dark matter is entirely comprised of a single particle whose antiparticle is itself. We also generalized the J factor expression into

$$J_{\mu,p}^\ell = \int_{\mu,p} d\Omega d\mu d^3v_1 d^3v_2 S_\ell \left(\frac{|\mathbf{v}_1 - \mathbf{v}_2|}{c} \right) \times f(\mathbf{r}(\mu), \mathbf{v}_1) f(\mathbf{r}(\mu), \mathbf{v}_2), \quad (4.11)$$

where $f(\mathbf{r}(\mu), \mathbf{v})$ is the distribution function of DM particles, which we integrate over the velocities of the two particles, \mathbf{v}_1 and \mathbf{v}_2 , and along the line of sight coordinate, μ . Note that taking $\ell = 0$ in Eq. (4.10) and Eq. (4.11) one will obtain the exact expressions from Eq. (4.3) and Eq. (4.5) respectively. Reason being that simply the integration over d^3v_1 and d^3v_2 in Eq. (4.11) becomes trivial and produces $\sim \rho^2$ term as written down in Eq. (4.5). Therefore, the thermal relic cross section can be identified with $a_\ell \equiv \langle \sigma v \rangle$. In the following, we will use these two interchangeably. The velocity dependent part of the cross section is defined to be

$$S_\ell(x) = x^{2\ell}, \quad (4.12)$$

where $\ell = 1$ (p -wave) or $\ell = 2$ (d -wave). This trivially reduces to unity for the s -wave case ($\ell = 0$).

The subscript p in Eq. (4.11) denotes the HEALPIX¹ pixel area. Given that we always calculate both the J and D factors on a HEALPIX grid, we also integrate across the pixel area p . Therefore, for consistency we can also integrate the original Eq. (4.6) for the D factor across the HEALPIX pixel and obtain

$$D_{p,\mu} = \int_{\mu,p} d\Omega d\mu \rho_{\text{DM}}(\mathbf{r}(\mu), \Omega), \quad (4.13)$$

In both the velocity independent ($\ell = 0$) and velocity dependent ($\ell = 1, 2$) J factor maps, we generate the templates on a higher resolution HEALPIX grid than that on which we perform the inference ($\text{nside} = 2048$ instead of $\text{nside} = 256$) and subsequently degrade them. Because of the nonlinear dependence of J on ρ , this allows for a more faithful representation of the density field than if J was initially calculated at $\text{nside} = 256$, which is especially important for the regions of the sky corresponding to halos produced in CSiBORG. We concluded the HEALPIX resolution of $\text{nside} = 2048$ was sufficient by comparing the total J factor to those with increasing HEALPIX resolution ($\text{nside} = 4096, 8192$) since the change in total J factor was at the subpercent level with increasing nside .

¹<http://healpix.sf.net>

4.2 Methods

In this section we detail how we construct the full-sky templates for DM annihilation and decay for all channels (*s*-, *p*- and *d*-wave), and how these are compared to the gamma-ray data to constrain the annihilation cross section and decay rate. In Section 4.2.1 we describe the constrained simulations used to generate these templates and in Section 4.2.3 we describe how the *J* and *D* factors are computed from the DM particles in these simulations. These templates are combined with those from Section 4.2.5 that describe non-DM contributions to the gamma-ray sky, and we compare these to the data using the likelihood model in Section 4.2.6.

4.2.1 Bayesian large-scale structure inference

In this work we study the expected gamma ray flux from dark matter annihilation and decay in extragalactic halos. To determine the masses and locations of these halos, we use the CSiBORG suite of 101 dark matter-only constrained simulations [183–185]. The initial conditions of these simulations produce present-day density fields which match the observed number densities of galaxies in the 2M++ galaxy compilation [23, 187] and are inferred using the BORG algorithm (see [19–22, 187]). This algorithm produces a Markov Chain of plausible initial conditions via the application of a Bayesian forward model and marginalises over galaxy bias parameters. The initial conditions are constrained in a box of length $677.7 h^{-1} \text{Mpc}$ with 256^3 voxels. For each CSiBORG simulation these initial conditions are augmented with white noise to a resolution of 2048^3 within $155 h^{-1} \text{Mpc}$ of the Milky Way (giving a mass resolution of $4.4 \times 10^9 M_{\odot}$) and run to $z = 0$ using RAMSES [194]. The watershed halofinder PHEW [195] is applied on-the-fly to the dark matter particles with the standard threshold of $200\rho_c$. We do not consider sub-halos and the minimum halo mass used ($M_{200\text{m}}$) is $4.4 \times 10^{11} M_{\odot}$. The resulting halo catalogues are publicly available at [186].

4.2.2 Smoothed density field contribution

The halo finding allows us to split CSiBORG particles into two types: those that belong to halos and those that do not. Since the *J* factor depends purely on the square of the density along the line of sight in case of *s*-wave, it is more sensitive to the small-scale matter distribution, and thus we must treat halos separately from the background density field in order to account for structures below the resolution of the CSiBORG simulations. This will not be an issue for the case of velocity dependent annihilation given the low velocity dispersion present inside the substructures. The *D* factor is less sensitive to these small scales, and thus we treat all particles equally in this case. *D* is computed using the procedure outlined in Section 4.2.2 below, where we use all particles. For the *s*-wave *J* factor we compute the above integral from Eq. (4.5), but only considering

nonhalo particles, and add this to the contribution from halos, which are treated as in Section 4.2.3. We plot the resulting ensemble mean J and D factor maps in Fig. 4.1.

For the case of velocity independent cross section and decay we also wish to determine the density of DM particles that do not belong to a halo on a regular Cartesian grid with $N_{\text{grid}} = 1024$ grid points per side. For the velocity dependent case, the velocity dispersion is not expected to be very high outside of formed halos and therefore this step is neglected.

In order to determine the DM density outside of halos, we adopt a procedure based on smooth particle hydrodynamics (SPH) [196] as described in [197] and outlined below. Using the SPH algorithm over, e.g., a cloud-in-cell (CIC) approach allows us to better capture the peaks of the matter density field, since the SPH kernel will adapt to the local density of matter, in contrast to the CIC approach which has a fixed kernel corresponding to a trilinear interpolation scheme. We compare the results of using a SPH kernel to a CIC algorithm in Section VI B.1 in [132].

First, we determine the number of particles, N_p , within the cell corresponding to each grid point (i, j, k) . We then define

$$N_X = \max(N_p, N_{\text{SPH}}), \quad (4.14)$$

where $N_{\text{SPH}} = 32$. The choice for this number of neighbours is partly motivated by the typical number of edges linking a node to its neighbours in a Delaunay tessellation. That number is approximately 16 for a Euclidean three-dimensional vector space [24, 198]. We pick a value twice as big as we intend the filter to have a larger reach than the first layer of neighbours. We then find the mass associated with this grid point by considering the nearest N_X particles to be

$$\tilde{m}_{ijk} = \frac{1}{R_{ijk}^3} \sum_{l=1}^{N_X-1} m_l W_l \mathcal{S} \left(\frac{d_l}{R_{ijk}} \right), \quad (4.15)$$

where R_{ijk} is half the distance to the farthest of the N_X particles from the grid point, m_l is the mass of particle l , which is at a distance d_l from the grid point, W_l is the weight for particle l ,

$$W_l = \left(\sum_{ijk} \frac{1}{R_{ijk}^3} \mathcal{S} \left(\frac{d_l}{R_{ijk}} \right) \right)^{-1}, \quad (4.16)$$

and the interpolating function, \mathcal{S} , is chosen to be

$$\mathcal{S}(x) = \begin{cases} 1 - \frac{3}{2}x + \frac{3}{4}x^3, & 0 \leq x < 1 \\ \frac{1}{4}(2-x)^3, & 1 \leq x < 2 \\ 0, & \text{otherwise.} \end{cases} \quad (4.17)$$

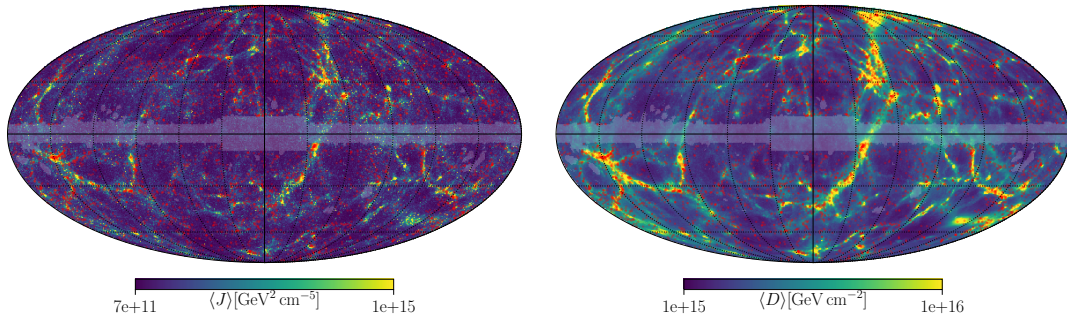


Figure 4.1: Mollweide projection in galactic coordinates of the ensemble mean s -wave ($\ell = 0$) J factor and the D factor over the CSiBORG realisations, alongside the brightest ~ 6000 galaxies from the 2M++ dataset (red points). One can see that the galaxy number density is higher in the regions of large J and D factor, i.e. at the peaks of the underlying DM distribution. We overplot the mask on completeness used in the BORG inference of the initial conditions [23, 187] (faded region near the galactic plane masked out).

If the spacing between grid points is Δr , then the density assigned to each site is

$$\tilde{\rho}_{ijk} = \frac{\tilde{m}_{ijk}}{\Delta r^3}. \quad (4.18)$$

To compute the J and D factors, we compute Eqs. (4.5) and (4.6), respectively, along the line of sight corresponding to each HEALPIX pixel at the chosen resolution. We integrate up to the edge of the simulated volume and perform trilinear interpolation of the density field onto the line of sight. The convergence of this approach was checked by increasing the resolution of the SPH kernel. The total assigned mass to the grid was consistent among all resolutions we tried ($N_{\text{grid}} = 256, 512, 1024$); therefore we opted for $N_{\text{grid}} = 1024$ for our final J and D factor calculations. Once more, it should be kept in mind that the nonhalo particle contribution to the J factor templates is only added in case of s -wave annihilation, while the p - and d -wave J factor calculation doesn't contain this contribution and the specific calculation is described in the following section. In case of the D factor however, we always include the contribution from the smoothed density field.

4.2.3 Computing the s -wave J factor

To include the contribution from particles inside halos, we use a custom extension of the CLUMPY package² [199–201]. In this section we review how the J factor is calculated in this package, as well as our assumptions for the halo density profiles.

²<https://clumpy.gitlab.io/CLUMPY>

We assume that all halos are spherically symmetric and that the total density profile, ρ_{tot} , can be described by a simple analytic form. For our fiducial case, we consider the three-parameter family of profiles [202, 203]

$$\rho_{\alpha\beta\gamma}(r) = \frac{2^{\frac{\beta-\gamma}{\alpha}} \rho_s}{(r/r_s)^\gamma (1 + (r/r_s)^\alpha)^{\frac{\beta-\gamma}{\alpha}}}, \quad (4.19)$$

where r_s is the scale radius, ρ_s is the density at r_s , and α describes the sharpness of the transition between the inner (γ) and outer (β) logarithmic slopes. For a NFW profile, $\alpha = 1$, $\beta = 3$, $\gamma = 1$. In our paper [132] we also considered the Einasto profile [182, 204], in order to test for systematics with regard to the choice of the profile. We refer the reader to the original paper for more details.

For the halo profile we calculate the parameters defining the profile itself from the total halo mass and corresponding concentration. The mass-concentration relation we use is shown in Eq. (4.29). Given that all of our halos are at very low redshift ($z \lesssim 0.05$) and almost all have masses within $M \in [10^{12}, 10^{15}] M_\odot$, the parametric relation is quite accurate. We use the COLOSSUS package throughout [205]. Baryons induce a potentially larger effect, which is, however, harder to model reliably. For more discuss this further in Section VI B.2 of [132]. We found that the constraints can be up to 80% tighter than for the corresponding NFW profile, given the flatness of the Einasto profiles towards the halo centre.

One would also expect that a halo contains a large number of ‘‘clumps’’ or subhalos, such that the true smooth component of the density profile is [206]

$$\rho_{\text{sm}}(r) = \rho_{\text{tot}}(r) - \langle \rho_{\text{subs}}(r) \rangle, \quad (4.20)$$

where $\langle \rho_{\text{subs}}(r) \rangle$ gives the average contribution from the substructure. If clump i has density profile ρ_{cl}^i , then it contributes to the total J factor value within a HEALPIX pixel, p , as

$$\begin{aligned} J_p &= \int_{p, \Delta_{\text{halo}}} \left(\rho_{\text{sm}}(\mu, \Omega) + \sum_i \rho_{\text{cl}}^i(\mu, \Omega) \right)^2 d\mu d\Omega \\ &= J_{\text{sm},p} + J_{\text{subs},p} + J_{\text{cross},p}, \end{aligned} \quad (4.21)$$

where

$$J_{\text{sm},p} = \int_{p, \Delta_{\text{halo}}} \rho_{\text{sm}}^2(\mu, \Omega) d\mu d\Omega, \quad (4.22)$$

$$J_{\text{subs},p} = \int_{p, \Delta_{\text{halo}}} \left(\sum_i \rho_{\text{cl}}^i(\mu, \Omega) \right)^2 d\mu d\Omega, \quad (4.23)$$

$$J_{\text{cross},p} = 2 \int_{p, \Delta_{\text{halo}}} \rho_{\text{sm}}(\mu, \Omega) \left(\sum_i \rho_{\text{cl}}^i(\mu, \Omega) \right) d\mu d\Omega, \quad (4.24)$$

and Δ_{halo} represents the intersection of the halo volume with the line of sight cone spanned by the pixel p . Our task is therefore to determine the distribution of clumps for a given halo and to calculate these integrals. Here, we readily use the solution provided by the CLUMPY package and describe it briefly below.

Assuming that a given halo has N_{tot} independent clumps, we factorise the distribution for the number of clumps with some mass M , concentration c , in some region $dV = d^3r$ as [200, 207]

$$\frac{dN}{dV dM dc} = N_{\text{tot}} \frac{d\mathcal{P}_V(r)}{dV} \frac{d\mathcal{P}_M(M)}{dM} \frac{d\mathcal{P}_c(M, c)}{dc}. \quad (4.25)$$

Since the clumps form before the host halos within Λ CDM, their spatial distribution will follow the host DM density profile. This has been shown to be a good assumption in simulations of galaxy-sized halos [182, 208]. Given the self-similar nature of collapse of Λ CDM halos, we extrapolate this conclusion to halos from our CSiBORG ensemble. We assume that the distribution of masses is a power law

$$\frac{d\mathcal{P}_M}{dM} \propto M^{-\alpha_M}, \quad (4.26)$$

in the range $M \in [10^{-6}M_{\odot}, 10^{-2}M_{\text{h}}]$ for a halo of mass M_{h} , where $\alpha_M = 1.9$ (see Section 2.3 of [209] and references therein). Again, the values are motivated by numerical simulations of Milky Way sized halos, which we extrapolate to bigger halos present in our forward model.

Besides modelling the uncertainty due to the spatial and mass distribution of substructure, the CLUMPY package also allows us to include the uncertainty in the mass-concentration relation. For the substructure component, we consider two cases for the concentration distribution. In the first case, we assume that the concentration of all substructure halos is a deterministic function of the mass

$$\frac{d\mathcal{P}_c}{dc} = \delta(c - \bar{c}(M)), \quad (4.27)$$

where we define δ as the Dirac-delta distribution. The second possibility that we follow is that the concentration is log normally distributed about this mean

$$\frac{d\mathcal{P}_c}{dc} = \frac{1}{\sqrt{2\pi}c\sigma_c(M)} \exp\left(-\frac{(\log c - \log(\bar{c}(M)))^2}{2\sigma_c^2}\right). \quad (4.28)$$

Motivated by [179], the substructure halos are assumed to have the following mass-concentration relation [210]

$$\bar{c}(M) = \sum_{j=0}^5 C_j \left[\ln\left(\frac{M}{h^{-1}M_{\odot}}\right) \right]^j, \quad (4.29)$$

where $C_j = (37.5153, -1.5093, 1.636 \times 10^{-2}, 3.66 \times 10^{-4}, -2.8927 \times 10^{-5}, 5.32 \times 10^{-7})$, with $\sigma_c = 0.0$; i.e., we assume all substructure halos of the same mass have the same concentration. In Section VI B.3 of [132] we also considered a nonzero value $\sigma_c = 0.2$ in Eq. (4.28), as motivated by [211–213] for a comparison. We found that the scatter in final value for the cross section constraint is of the order of 1.4%.

Given that we do not resolve substructures of all the halos present with our simulation, we assume the resulting distribution for J_{sm} and J_{subs} to be a Gaussian. Therefore, we only need to find the mean and variance of these contributions in each HEALPIX pixel. We define the 1-clump luminosity to be

$$L(M, c) \equiv \int_{V_{\text{subhalo}}} \rho_{\text{cl}}^2(r; M, c) d^3r, \quad (4.30)$$

and its moments as

$$\langle L^n \rangle \equiv \int_{M_1}^{M_2} \frac{d\mathcal{P}_M}{dM} \int \frac{d\mathcal{P}_c}{dc} L^n dc dM, \quad (4.31)$$

for a given mass range of clumps $[M_1, M_2]$, while the mean contribution of $J_{\text{subs},p}$ from this volume along the line of sight is

$$\langle J_{\text{subs},p} \rangle = N_{\text{tot}} \int_{p, \mathcal{V}_\mu} \frac{d\mathcal{P}_V}{dV} \mu^{-2} dV \langle L \rangle, \quad (4.32)$$

with N_{tot} representing the total number of clumps within the selected mass range $[M_1, M_2]$. Note that we assume that the clumps are nonoverlapping, such that the cross terms in Eq. (4.23) can be neglected. For more details on how these quantities are defined we refer the reader to the CLUMPY related publications [199–201].

Note that since we are assuming unresolved substructures for our CSiBORG extragalactic halos, we are integrating over the total subhalo volume V_{subhalo} for the subhalo luminosity. Furthermore, since there can be many subhalos present within the line of sight determined by the given HEALPIX pixel we are also accounting for the span of the host halo along this line of sight through $\mathcal{V}_\mu \equiv [\mu_{\text{min}}, \mu_{\text{max}}]$, with μ_{min} and μ_{max} being the closest and farthest points of the host halo along this line of sight. Since these integrals do not have a closed form for general DM profiles, we evaluate all numerically.

Given that the mean of some power of the distance from the observer along the line of sight, μ , to a clump which falls inside the HEALPIX pixel p is

$$\langle \mu_p^n \rangle = \int_{p, \mathcal{V}_\mu} \mu^{n+2} \frac{d\mathcal{P}_V}{dV} d\mu d\Omega, \quad (4.33)$$

we can write the variance on J_{subs} as

$$\sigma_{J_{\text{subs},p}}^2 = \langle L^2 \rangle \langle \mu_p^{-4} \rangle - \langle L \rangle^2 \langle \mu_p^{-2} \rangle^2, \quad (4.34)$$

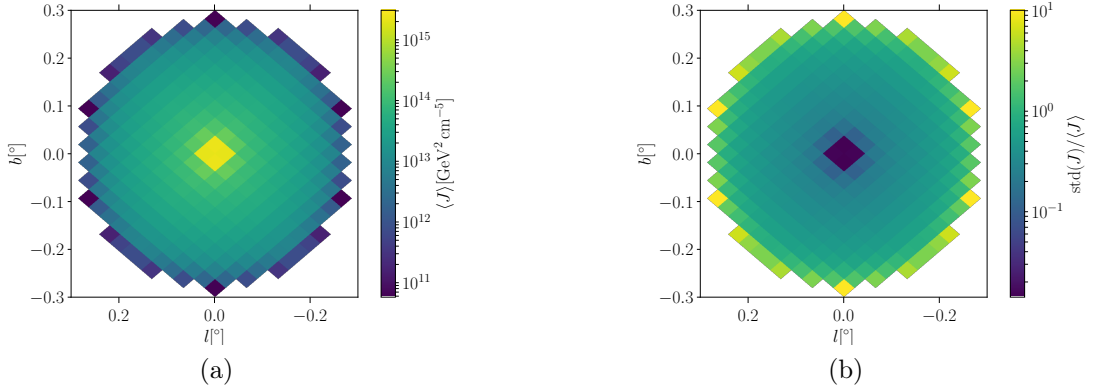


Figure 4.2: (a) The s -wave ($\ell = 0$) J factor of a typical NFW halo within CSiBORG ($M_h \approx 5 \times 10^{13} M_\odot$) and (b) the corresponding relative fluctuations in the J factor due to the substructure contribution. The quantities $\langle J \rangle$ and $\text{std}(J)$ are calculated according to Eqs. (4.39) and (4.40) respectively. As can be seen, the relative size of fluctuations in the J factor grows toward the outskirts. Qualitatively similar features are observed if we assume an Einasto profile. Note that here we placed the halo at the centre of the HEALPIX grid for numerical convenience.

since L and μ are independent. For the cross term, $J_{\text{cross},p}$, we use that its mean is

$$\langle J_{\text{cross},p} \rangle = 2 \int_{p, \mathcal{V}_\mu} \rho_{\text{sm}} \langle \rho_{\text{subs}} \rangle d\mu d\Omega, \quad (4.35)$$

while its variance can be computed as

$$\sigma_{J_{\text{cross},p}}^2 = 4 \int_{p, \mathcal{V}_\mu} \rho_{\text{sm}}^2(\mu, \Omega) \sigma_{\text{subs}}^2(\mu, \Omega) d\mu d\Omega, \quad (4.36)$$

with

$$\sigma_{\text{subs},p}^2 \equiv \sigma_{\text{subs}}^2(\mu, \Omega) = \langle \rho_{\text{subs}}^2(\mu, \Omega) \rangle - \langle \rho_{\text{subs}}(\mu, \Omega) \rangle^2, \quad (4.37)$$

$$\begin{aligned} \langle \rho_{\text{subs}}(\mu, \Delta\Omega) \rangle &= \int_{\Delta\Omega, \mathcal{V}_\mu} \frac{d\mathcal{P}_V}{dV} \int_{\mathcal{V}_M} \frac{d\mathcal{P}_M}{dM} \int_{\mathcal{V}_{c(M)}} \frac{d\mathcal{P}_c}{dc} \\ &\times \rho_{\text{subs}}(\mu, \Omega; M, c(M)) d\mu d\Omega dM dc, \end{aligned} \quad (4.38)$$

with \mathcal{V}_M and $\mathcal{V}_{c(M)}$ representing the mass and corresponding concentration ranges for the subhalo distribution respectively.

We decide to include only $\sigma_{J_{\text{subs},p}}$ as it is the dominant source of uncertainty. This can intuitively be understood from the Eq. (4.36). We note that the integrand is negligible both in the outskirts of the host halo, since $\rho_{\text{sm}} \sim 0$ and σ_{subs}^2 remains finite, and in

the very centre of the host halo, since $\sigma_{\text{subs},p} \sim 0$. Therefore, $J_{\text{cross},p}$ contributes only at a very limited range of scales. Furthermore, for our halos, there is a clear hierarchy between the cross and subs term $J_{\text{cross},p} \lesssim 0.01 - 0.1 J_{\text{subs},p}$; therefore, we focus only on the $\sigma_{J_{\text{subs},p}}$ as the dominant source of uncertainty of the J factor due to substructure.

We hence write that the distribution followed by the J factor for a given halo in a given pixel is given by a Gaussian with mean

$$\langle J_p \rangle = J_{\text{sm},p} + \langle J_{\text{subs},p} \rangle + \langle J_{\text{cross},p} \rangle, \quad (4.39)$$

and variance

$$\sigma_p^2 = \sigma_{J_{\text{subs},p}}^2 + \sigma_{J_{\text{cross},p}}^2 \approx \sigma_{J_{\text{subs},p}}^2. \quad (4.40)$$

We calculate these quantities for all CSiBORG realisations. To distinguish between these, we introduce a second index, j , to label the simulation, i.e. $\langle J_{pj} \rangle$ is $\langle J_p \rangle$ for CSiBORG simulation j , and likewise σ_{pj} is σ_p for the same simulation.

In Fig. 4.2, we show the result of the model for a typical halo within CSiBORG with a mass of $M_h \approx 5 \times 10^{13} M_\odot$. We see that the effects from the term in $\sigma_{J_{\text{subs}}}^2$ cannot be neglected, especially in the outskirts of the halo. In the very centre, where the structure of the halo is dominated by the smooth component, the fluctuations in the J factor due to the substructure are negligible, amounting to only few percent, while in the outer parts these fluctuations become more important. This is an expected result given that the boost in J factor due to substructure becomes more important in the outer edges, where the smooth component, $J_{\text{sm},p}$, is subdominant with respect to the substructure J factor, $J_{\text{subs},p}$. This behaviour is identical for an Einasto profile.

Besides this, we also include the contributions of sub-subclumps to the J factor of halos, using one additional level of substructure, which is the default setting of the CLUMPY code. Because of the increased computational cost, we considered a two-level substructure contribution for our halos for only one CSiBORG realisation. Including additional substructure levels will result in an overall change in the J factor of less than $\sim 5\%$ [200], and ignoring such levels will make our constraints on $\langle \sigma v \rangle$ conservative since this will systematically underestimate the J factor.

In conclusion, to obtain the total all-sky J factor, we combine the line-of-sight calculation for the density field obtained from particles outside of halos detected within CSiBORG realisations with the component coming from the halo particles of the CSiBORG by treating the halos as presented in this section, utilising the CLUMPY code. This final template is used in the inference pipeline. We discussed the numerical convergence of these calculations in Section VI B.1 of [132].

4.2.4 Computing the p - and d -wave J factor

For p -wave annihilation, the J factor can be re-written as

$$\begin{aligned} J_{\mu,p}^{(\ell=1)} &= 2 \int_{\mu,p} d\Omega d\mu \rho^2(\mathbf{r}(\mu)) \left\langle \frac{\mathbf{v}^2}{c^2} \right\rangle(\mathbf{r}(\mu)) \\ &= \frac{2}{c^2} \int_{\mu,p} d\Omega d\mu \rho^2(\mathbf{r}(\mu)) (\sigma_r^2(\mathbf{r}(\mu)) + 2\sigma_\theta^2(\mathbf{r}(\mu))), \end{aligned} \quad (4.41)$$

where $\rho(\mathbf{r}(\mu))$ is the dark matter density at the position μ along the line of sight and σ_r and σ_θ are the corresponding radial and tangential velocity dispersion, respectively. We keep the explicit dependence on the line of sight position in order to emphasize that the J factor needs to be evaluated with respect to the observer. However, since the total J factor within a pixel is a scalar quantity, one can also first evaluate it within the corresponding volume with respect to the halo centre and then make the appropriate projection. This is precisely how the CLUMPY package – which we utilise here – works (see [132, 199–201]). Now, to obtain the dispersion for a given halo, we assume spherical symmetry, and therefore must solve the Jeans equation

$$\frac{1}{\rho} \frac{d}{dr} (\rho \sigma_r^2) + 2\beta \frac{\sigma_r^2}{r} = -\frac{d\Phi}{dr}, \quad (4.42)$$

where we introduce an anisotropy parameter

$$\beta(r) = 1 - \frac{\sigma_\theta^2(r)}{\sigma_r^2(r)}, \quad (4.43)$$

and Φ is the gravitational potential of the halo.

We assume that all our halos have Navarro-Frenk-White (NFW) [188] density profiles, taking $\alpha = \beta = \gamma = 1$ in Eq. (4.19)

$$\rho(r) = \frac{\rho_s}{(r/r_s)(1+r/r_s)^2}. \quad (4.44)$$

Here, we also use the mean mass-concentration relation of [210, 214] and assume zero scatter in this relation. Since the J factor traces the square of the density, one may expect that subhalos or clumps may provide a significant contribution to the p - or d -wave J factor and should also be included. Although this is true for s -wave annihilation, the small velocity dispersion of these clumps means that the overall signal is dominated by the smooth component and we therefore ignore unresolved substructure. This approximation has been validated by explicitly calculating the p - and d -wave J factor for halos from hydrodynamical simulations [189], although we note that this conclusion does not hold for the case of Sommerfeld enhanced annihilation (see for example [190, 191]). We do not consider this case here and leave it for future work.

Defining the dimensionless radial coordinate $s \equiv r/r_{\text{vir}}$, one finds that the radial velocity dispersion for circular orbits is (see Eq. (14) in [215])

$$\begin{aligned} \frac{\sigma_r^2}{V_{\text{vir}}^2}(s) \stackrel{\beta=0}{=} & \frac{c^2 s (1+cs)^2}{2(\ln(1+c) - c/(1+c))} \left[\pi^2 + 6 \text{Li}_2(-cs) \right. \\ & - \ln(cs) + \left(1 + \frac{1}{c^2 s^2} - \frac{4}{cs} - \frac{2}{1+cs} \right) \times \ln(1+cs) \\ & \left. + 3 \ln^2(1+cs) - \frac{1}{cs} - \frac{6}{1+cs} - \frac{1}{(1+cs)^2} \right], \end{aligned} \quad (4.45)$$

where V_{vir} is the virial velocity, c the halo concentration and Li_2 is the dilogarithm function. Similarly, for the extreme cases of a constant anisotropy parameter $\beta = 0.5$ and $\beta = 1.0$ one finds (see Eq. (15) and (16) from [215])

$$\begin{aligned} \frac{\sigma_r^2}{V_{\text{vir}}^2}(s) \stackrel{\beta=0.5}{=} & \frac{c(1+cs)^2}{(\ln(1+c) - c/(1+c))} \left[-\frac{\pi^2}{3} - 2 \text{Li}_2(-cs) \right. \\ & + \frac{2}{1+cs} + \frac{\ln(1+cs)}{cs} + \frac{\ln(1+cs)}{1+cs} \\ & \left. - \ln^2(1+cs) - \frac{1}{2(1+cs)^2} \right], \quad (4.46) \\ \frac{\sigma_r^2}{V_{\text{vir}}^2}(s) \stackrel{\beta=1}{=} & \frac{(1+cs)^2}{s(\ln(1+c) - c/(1+c))} \left[\frac{\pi^2}{6} + \text{Li}_2(-cs) \right. \\ & \left. - \frac{1}{1+cs} - \frac{\ln(1+cs)}{1+cs} + \frac{\ln^2(1+cs)}{2} \right]. \end{aligned}$$

In Fig. 4.3 we show how these different velocity dispersion profiles affect the angular dependence of the normalised p -wave J factor (labelled as $\tilde{J}^{(\ell=1)}$) for a halo of mass $M = 1.2 \times 10^{15} M_{\odot}$ (one of the most massive halos in CSiBORG simulation 9844) at a distance of ~ 100 Mpc from the observer. The quantity \tilde{J} we calculate as

$$\tilde{J}^{\ell}(\theta) = \frac{J^{\ell}(\theta)}{\int d\Omega J^{\ell}(\theta)}, \quad (4.47)$$

where θ is the angular distance from the halo centre as seen on the sky. We perform this calculation from close to the halo centre, $\theta \sim 0.01^{\circ}$, all the way up to the two times the virial radius of the halo. Similar results for the case of the Milky Way halo (modelled as a NFW profile) were obtained through the Eddington inversion method in [216]. We show the results for p -wave with different values of the anisotropy parameter β (Eqs. (4.45) and (4.46)) and d -wave calculations. The $\beta = 0.5$ (dashed) and $\beta = 1.0$ (dotted-dashed) cases are more strongly peaked at the centre than the corresponding $\beta = 0$ case, reflecting the behaviour of the velocity dispersion profiles (see also Fig. 1

in [215]). Furthermore, for $\beta = 1.0$, \tilde{J} diverges at the centre and hence needs to be regularised (see Eq. (4.75)). The d -wave \tilde{J} shows qualitatively similar behaviour as the corresponding p -wave case (thick line). For more details on the d -wave calculation, see Eq. (4.48) and discussion below it as well as Section 4.7.

It should be noted that, for a more realistic scenario, the anisotropy parameter should be taken as a radially varying function (see [217–219]). We do not consider these models here, given that the three extreme scenarios for the anisotropy parameter we pick should bracket reality. In our fiducial analysis we choose $\beta = 0$, however, and show in Section 4.5.1 that this is a conservative choice. The corresponding J factor calculation then simply consists of substituting Eq. (4.45) into Eq. (4.41), noting that in this case $\sigma_\theta = \sigma_r$, and integrating the final expression for all our halos. We do this numerically using the CLUMPY package at a HEALPIX resolution of `nside` = 2048, which is then subsequently degraded to the coarser resolution (`nside` = 256) at which we perform the inference. This allows a more faithful calculation of the J factor than performing the calculation at `nside` = 256, since the dependence of J on ρ is non-linear.

For d -wave annihilation the corresponding expression for the J factor takes the form (see also [220])

$$J_{\mu,p}^{(\ell=2)} = \int_{\mu,p} d\Omega d\mu \rho^2(\mathbf{r}) \left(2 \left\langle \frac{\mathbf{v}^4}{c^4} \right\rangle(\mathbf{r}) + \frac{10}{3} \left\langle \frac{\mathbf{v}^2}{c^2} \right\rangle^2(\mathbf{r}) \right). \quad (4.48)$$

We see that the second term appearing can be computed from a power of the velocity dispersion (as in the p -wave case; see Eq. (4.41)), but the first term requires the evaluation of the fourth moment of the velocity distribution. In general, this would require for solving the Boltzmann hierarchy up to the fourth moment. It will not be possible to do so, however, since this system of equations is not closed for a general self-gravitating stellar system. In order to solve the Jeans system of differential equations, one must impose additional constraints. This can be done by requiring the stellar system has certain symmetries. One of the possibilities is to assume an ergodic distribution function ($\beta = 0$, see Section 4 - Box 4.3 of [221]). This then allows to write down the following equation, relating $\langle v_r^2 \rangle$ and $\langle v_r^4 \rangle$

$$\frac{d}{dr} (\rho \langle v_r^4 \rangle) \stackrel{\beta=0}{=} -3\rho \langle v_r^2 \rangle \frac{d\Phi}{dr}, \quad (4.49)$$

where by v_r we have denoted the projection of the velocity to the radial direction (fixing the coordinate system to the halo centre of mass). In addition to calculating $\langle v_r^4 \rangle$, we need to also calculate the other projected moments as well as the cross-terms in order to obtain the total fourth moment of the velocity, as is necessary for evaluating the Eq. (4.48). This is straightforward, due to the ergodicity of the distribution function, as shown in Section 4.7. The final expression for the fourth velocity moment is then given by

$$\langle \mathbf{v}^4 \rangle \stackrel{\beta=0}{=} 5 \langle v_r^4 \rangle. \quad (4.50)$$

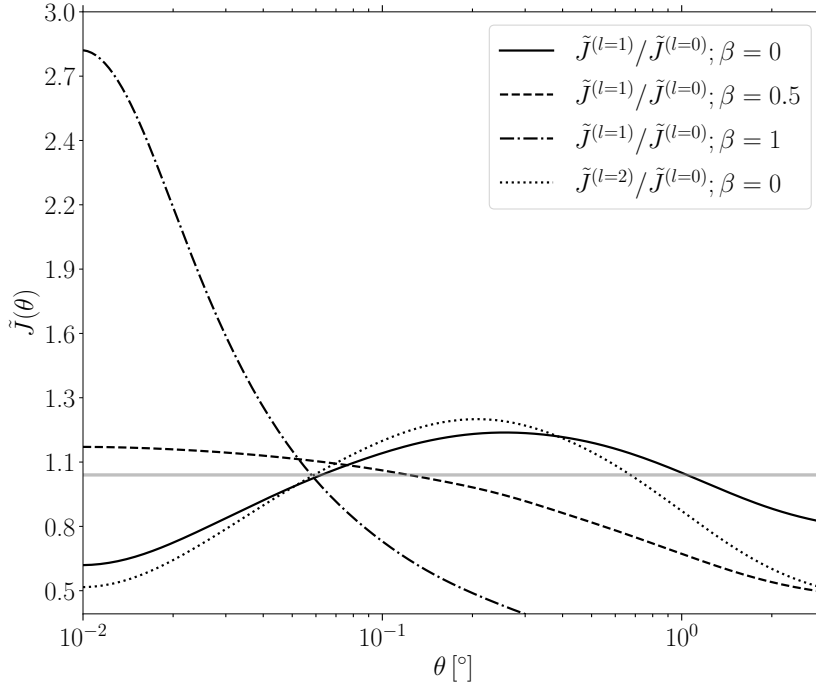


Figure 4.3: The angular distribution of the normalised J factor, \tilde{J} (Eq. 4.47), as a function of angle on the sky from the halo centre, θ , for p -wave ($\ell = 1$) and d -wave ($\ell = 2$) annihilation channels. The results are shown relative to the s -wave ($\ell = 0$) result and for different choices of anisotropy parameter, β . For this example, we used one of the most massive halos from the CSiBORG 9844 realisation with $M = 1.2 \times 10^{15} M_{\odot}$ at a distance of 96 Mpc from the observer.

This, and the solution for $\langle v^2 \rangle$ from Eq. (4.45), are then enough to evaluate the total J factor for d -wave annihilation. See Section 4.7 for more details.

4.2.5 Gamma ray data

We compare the dark matter annihilation templates to the gamma ray observations from *Fermi* Large Area Telescope between mission weeks 9 and 634. We select photons in the upper quartile of angular resolution (PS3) and event class SOURCEVETO in the energy range 500 MeV – 50 GeV. The maximum zenith angle is chosen to be 90° . These are binned spatially onto HEALPIX [222, 223] maps with resolution `nside` = 256 and into 9 logarithmically spaced energy bins. The processing of the photon files is performed

using the Fermi Tools³ and FERMIPY [224]. We also use this software to convolve our templates with the *Fermi* point spread function and to obtain the exposure maps. To reduce the impact of galactic emission and the GCE, we mask the region with galactic latitude $|\lambda| < 30^\circ$.

Although some of the photons detected by *Fermi* could have been produced by dark matter annihilation, there are other sources which will also contribute, and thus it is important to incorporate these in the analysis. We therefore consider three contributions: point sources (psc), an isotropic background (iso) and the Milky Way (gal) (see [132] for more details). Each contribution is modelled with a fixed spatial template for energy bin i , $\{T_i^t(\hat{r}) : t \in \{\text{iso}, \text{gal}, \text{psc}\}\}$, where we assign a different normalisation, A_i^t , for each energy bin and template. We jointly infer these parameters with the amplitude of the J or D factor template. In case of p - or d -wave we only consider the J factor template contribution. The point source template is comprised of all extended and point sources in the Large Area Telescope 12-year Source Catalog (4FGL-DR3)⁴ and the galactic component is modelled using the templates described in [225]. The scalings $\{A_i^t\}$ are defined such that they all equal unity if the *Fermi* data is perfectly described by these templates, however jointly inferring these parameters will allow us to capture any imperfections in the spectral modelling. We thus require an initial guess for the isotropic component's spectral shape, for which we use the *Fermi* Isotropic Spectral Template⁵. All templates are convolved with the *Fermi* point spread function using the Fermi Tools.

4.2.6 Likelihood model

Instead of directly constraining the DM annihilation or decay parameters, we split the inference into two parts. First, we assume that there is a contribution to the gamma-ray sky which is proportional to the J or D factor, i.e. for a given CSiBORG simulation j , the flux in energy bin i and pixel p , Φ_{ipj} , has terms

$$\Phi_{ipj} \supset \left(A_i^J \frac{J_{pj}}{J_0} + A_i^D \frac{D_{pj}}{D_0} \right) \mathcal{A}_p \Delta E_i, \quad (4.51)$$

where ΔE_i is the width of the bin, \mathcal{A}_p is the area of the pixel in steradians and J_0 and D_0 set the units. We choose $J_0 = 10^{13} \text{ GeV}^2 \text{ cm}^{-5}$ and $D_0 = 10^{13} \text{ GeV cm}^{-2}$. We fit for the total flux of such a contribution in each energy bin to obtain a spectrum. In the second half of the inference, we fit this spectrum to a series of models (including DM annihilation and decay) in an attempt to determine the origin of such a signal.

This method has two main advantages. First, we can consider each energy bin and CSiBORG simulation separately in the first half of the inference. Although this involves

³<https://fermi.gsfc.nasa.gov/ssc/data/analysis/software/>

⁴<https://heasarc.gsfc.nasa.gov/W3Browse/fermi/fermilpsc.html>

⁵<https://fermi.gsfc.nasa.gov/ssc/data/access/lat/BackgroundModels.html>

initially computing 909 Markov Chain Monte Carlo (MCMC) chains (one for each energy bin and for each CSiBORG simulation), since the problem is embarrassingly parallelisable and because we only need to infer four or five parameters for each chain (A_i^J and/or A_i^D , A_i^{iso} , A_i^{gal} , A_i^{psc}) compared to ~ 30 if we combined the energy bins, we find that this approach is computationally more efficient. Second, by remaining agnostic to the origin of A_i^J or A_i^D until the second step, we are able to more easily determine which energy bins drive our constraints. Hence, it becomes simpler to compare different models since we do not need to rerun the map-level inference every time that we wish to change the DM particle mass or decay channel (which is more computationally expensive).

Inferring the spectrum

We assume that photon counts in energy bin i from the J and D factor contributions, as well as each of the contributions described in Section 4.2.5 is Poisson distributed. The variation of the mean of the latter with sky position, \hat{r} , and energy is described by the known templates $\{T_i^t(\hat{r})\}$, where t labels the templates. For pixel p and CSiBORG realisation j , we then define

$$J_{ipj}^\ell \equiv \frac{J_{pj}^\ell}{J_0} \Delta E_i, \quad D_{ipj} \equiv \frac{D_{pj}}{D_0} \Delta E_i, \quad T_{ip}^t \equiv \int_p T_i^t(\hat{r}) d\Omega, \quad (4.52)$$

such that the mean number counts in pixel p and energy bin i is predicted to be

$$\lambda_{ipj}^\ell = \mathcal{F}_{ip} \times \left(A_i^{J^\ell} J_{ipj}^\ell + A_i^D D_{ipj} \delta_D^{\ell,0} + \sum_t A_i^t T_{ip}^t \right), \quad (4.53)$$

and we have multiplied our templates by the *Fermi* exposure, \mathcal{F}_{ip} , which describes the angular variation of the sensitivity of *Fermi*. The index ℓ distinguishes between the s - ($\ell = 0$), p - ($\ell = 1$) and d -wave ($\ell = 2$) cases. We emphasize that the D factor contribution is only accounted for when $\ell = 0$. The calculation of Eq. (4.53) is performed using the Fermi Tools, where we also convolve all templates with the point spread function. The likelihood of observing n_{ip} counts in pixel p and energy bin i given the mean λ_{ipj} is

$$\mathcal{L}(n_{ip} | \lambda_{ipj}^\ell) = \frac{(\lambda_{ipj}^\ell)^{n_{ip}} \exp(-\lambda_{ipj}^\ell)}{n_{ip}!}. \quad (4.54)$$

For the case of p - and d -wave annihilation analysis, this will be our fiducial likelihood. However, as discussed in Section 4.2.3, we do not know the exact DM distribution for a given CSiBORG simulation due to the unresolved substructure in halos. This is especially relevant for the case of s -wave J factor template. We did not include this uncertainty in Eq. (4.54). The uncertainty on the substructure contribution to the s -wave J factor is modelled as a truncated Gaussian. This choice allows us to marginalise analytically over

the substructure uncertainty. Therefore, for our fiducial s -wave analysis our likelihood can be obtained as

$$\begin{aligned} \mathcal{L} \left(\lambda_{ipj}^{(\ell=0)} | A_i^{J^{(\ell=0)}}, A_i^D, \{A_i^t\}, j \right) &= \frac{1}{\sigma_{ipj}} \sqrt{\frac{2}{\pi}} \left(1 + \operatorname{erf} \left(\frac{\xi_{ipj}}{\sigma_{ipj} \sqrt{2}} \right) \right)^{-1} \\ &\times \exp \left(-\frac{\left(\lambda_{ipj}^{(\ell=0)} - \xi_{ipj} \right)^2}{2\sigma_{ipj}^2} \right), \end{aligned} \quad (4.55)$$

for $\lambda_{ipj}^{(\ell=0)} > 0$, and zero otherwise, where

$$\xi_{ipj} = \mathcal{F}_{ip} \times \left(A_i^{J^{(\ell=0)}} \langle J_{ipj} \rangle + A_i^D \langle D_{ipj} \rangle + \sum_t A_i^t T_{ip}^t \right), \quad (4.56)$$

and

$$\sigma_{ipj} = A_i^{J^{(\ell=0)}} \frac{\sigma_{pj}}{J_0} \Delta E_i. \quad (4.57)$$

Therefore, the likelihood for observing n_{ip} counts in pixel p and energy bin i for the case of s -wave templates is

$$\begin{aligned} \mathcal{L} \left(n_{ip} | A_i^{J^{(\ell=0)}}, A_i^D, \{A_i^t\}, j \right) &= \int \mathcal{L}_{ipj} \left(n_{ip} | \lambda_{ipj}^{(\ell=0)} \right) \mathcal{L} \left(\lambda_{ipj}^{(\ell=0)} | A_i^J, A_i^D, \{A_i^t\}, j \right) d\lambda_{ipj}^{(\ell=0)} \\ &= \sqrt{\frac{2^n \sigma_{ipj}^{n_{ip}}}{\pi n_{ip}!}} \exp \left(-\frac{\xi_{ipj}^2}{2\sigma_{ipj}^2} \right) \left(1 + \operatorname{erf} \left(\frac{\xi_{ipj}}{\sigma_{ipj} \sqrt{2}} \right) \right)^{-1} \\ &\times \left(\Gamma \left(\frac{n_{ip} + 1}{2} \right) {}_1F_1 \left(\frac{n_{ip} + 1}{2}; \frac{1}{2}; \frac{(\xi_{ipj} - \sigma_{ipj}^2)^2}{2\sigma_{ipj}^2} \right) \right. \\ &+ \sqrt{2} \frac{(\xi_{ipj} - \sigma_{ipj}^2)}{\sigma_{ipj}} \Gamma \left(\frac{n_{ip}}{2} + 1 \right) \\ &\left. \times {}_1F_1 \left(\frac{n_{ip} + 2}{2}; \frac{3}{2}; \frac{(\xi_{ipj} - \sigma_{ipj}^2)^2}{2\sigma_{ipj}^2} \right) \right), \end{aligned} \quad (4.58)$$

where ${}_1F_1$ is the confluent hypergeometric function of the first kind. We described how we implement this likelihood numerically in Appendix A in [132].

Assuming that each pixel is independent, the likelihood for the observed data in energy bin i , \mathcal{D}_i , is

$$\mathcal{L} \left(\mathcal{D}_i | A_i^J, A_i^D, \{A_i^t\}, j \right) = \prod_p \mathcal{L} \left(n_{ip} | A_i^J, A_i^D, \{A_i^t\}, j \right). \quad (4.59)$$

Using the priors, P , given in Table 4.1, we apply Bayes' identity

$$\begin{aligned} \mathcal{L} \left(A_i^{J^\ell}, A_i^D, \{A_i^t\}, j | \mathcal{D}_i \right) \\ = \frac{\mathcal{L} \left(\mathcal{D}_i | A_i^{J^\ell}, A_i^D, \{A_i^t\}, j \right) P \left(A_i^{J^\ell} \right) P \left(A_i^D \right) P \left(\{A_i^t\} \right) P \left(j \right)}{\mathcal{Z} \left(\mathcal{D}_i \right)}, \end{aligned} \quad (4.60)$$

where

$$P \left(\{A_i^t\} \right) \equiv \prod_t P \left(A_i^t \right), \quad (4.61)$$

to obtain the posterior, $\mathcal{P} \left(A_i^{J^\ell}, A_i^D, \{A_i^t\}, j | \mathcal{D}_i \right)$, where $\mathcal{Z} \left(\mathcal{D}_i \right)$ is the evidence and where we consider each energy bin and CSiBORG simulation separately. We use the EMCEE sampler [87] and terminate the chain when the estimate of the autocorrelation length changes by less than 1 percent per iteration and the chain is at least 100 autocorrelation lengths long in all of the parameters.

We apply a Monte Carlo estimate to the likelihood of the CSiBORG samples, such that the one-dimensional posterior for the amplitude A_i^X (X being J^ℓ or D) is

$$\mathcal{L} \left(A_i^X | \mathcal{D}_i \right) = \frac{1}{N_{\text{sim}}} \sum_j \int d\{A_i^t\} dA_i^Y \mathcal{L} \left(A_i^{J^\ell}, A_i^D, \{A_i^t\}, j | \mathcal{D}_i \right), \quad (4.62)$$

where $Y = J^\ell$ if $X = D$ and vice versa. Again, the D factor is only considered in case of s -wave analysis ($\ell = 0$). In practice, we compute the average over CSiBORG realisations by first fitting the one-dimensional posteriors $\mathcal{L} \left(A_i^X | \mathcal{D}_i, j \right)$ with a spline using the `GetDist` package [226] and then computing the mean of the resulting functions. This is equivalent to concatenating the Markov Chains if each chain had the same length.

Constraining DM parameters

We now have a posterior, $\mathcal{L} \left(A_i^{J^\ell} | \mathcal{D}_i \right)$, describing the gamma-ray spectrum from sources that have the same spatial distribution as the J^ℓ factor. We wish to fit this spectrum to a model, $f_i(\boldsymbol{\theta})$, for these sources and infer the model parameters $\boldsymbol{\theta}$. We assume that our model is deterministic, such that

$$\mathcal{L} \left(A_i^{J^\ell} | \boldsymbol{\theta} \right) = \delta \left(A_i^{J^\ell} - f_i(\boldsymbol{\theta}) \right), \quad (4.63)$$

Parameter	Prior
$m_\chi / \text{GeV}/c^2$	[2, 500]
$\langle\sigma v\rangle / 10^{-26} \text{cm}^3\text{s}^{-1}$	[0, 10^3]
$\Gamma / 10^{-30} \text{s}^{-1}$	[0, 10^3]
A_i^t	[0.5, 1.5]
$A_i^J / 10^{-16} \text{cm}^{-2}\text{s}^{-1}\text{MeV}^{-1}$	$[0, (300 \text{GeV}/E_i)^2]$
$A_i^D / 10^{-16} \text{cm}^{-2}\text{s}^{-1}\text{MeV}^{-1}$	$[0, 0.5 \times (300 \text{GeV}/E_i)^2]$

Table 4.1: Priors on DM properties and template amplitudes ($A_i^t \in \{A_i^{\text{gal}}, A_i^{\text{iso}}, A_i^{\text{psc}}\}$), as defined in the text. All priors are uniform in the range given, except from the DM particle mass, m_χ , since we constrain the cross section, $\langle\sigma v\rangle$, and decay rate, Γ , at fixed m_χ . The priors on A_i^J and A_i^D depend on the minimum energy of the energy bin, E_i , although in all cases the prior is much wider than the posterior. For DM decay we also ensure that m_χ is at least twice the rest mass of the final decay products, and for annihilation this limit is equal to the rest mass of the standard model particle.

and therefore we obtain the likelihood for the observed gamma-ray sky by incorporating Eq. (4.62),

$$\begin{aligned} \mathcal{L}(\mathcal{D}_i|\boldsymbol{\theta}) &= \int dA_i^{J^\ell} \mathcal{L}(\mathcal{D}_i|A_i^{J^\ell}) \mathcal{L}(A_i^{J^\ell}|\boldsymbol{\theta}) \\ &= \int dA_i^{J^\ell} \frac{\mathcal{L}(A_i^{J^\ell}|\mathcal{D}_i) \mathcal{Z}(\mathcal{D}_i)}{P(A_i^{J^\ell})} \delta(A_i^{J^\ell} - f_i(\boldsymbol{\theta})). \end{aligned} \quad (4.64)$$

We assume that all energy bins are independent such that the likelihood of $\boldsymbol{\theta}$ given the full dataset \mathcal{D} is

$$\mathcal{L}(\mathcal{D}|\boldsymbol{\theta}) = \prod_i \mathcal{L}(\mathcal{D}_i|\boldsymbol{\theta}), \quad (4.65)$$

and so with a final application of Bayes' identity we obtain the posterior for our model parameters

$$\mathcal{L}(\boldsymbol{\theta}|\mathcal{D}) = \frac{\mathcal{L}(\mathcal{D}|\boldsymbol{\theta}) P(\boldsymbol{\theta})}{\mathcal{Z}(\mathcal{D})}. \quad (4.66)$$

If $f_i(\boldsymbol{\theta})$ comprises exclusively of DM annihilation, then, at fixed DM mass and annihilation channel, the transformation from the posterior distribution of $A_i^{J^\ell}$ to that of a_ℓ is trivial. For more complicated models (i.e., where $\boldsymbol{\theta}$ consists of more than one parameter), we again calculate the posterior on $\boldsymbol{\theta}$ using the EMCEE and NUMPYRO packages.

For DM annihilation and decay, prompt production, decays, hadronisation and radiative processes associated with the resulting standard model products produce a variety of stable species, including gamma rays. For a given channel, one must know the energy

spectrum of the intermediate standard model particles and the resulting branching ratios and energies of the subsequently produced particles. One then has, for each channel, a model for the spectrum of gamma rays as a function of DM particle mass and annihilation cross-section or decay rate.

In this work we utilise the pre-computed spectra provided by the *Fermi* collaboration⁶ which are calculated as described by [227]. Since we are considering non-relativistic s -wave annihilation in this work, one can view the annihilation of two DM particles of mass m_χ as equivalent to the decay of a single particle of mass $2m_\chi$ [228]. Hence, we obtain the spectrum for decay from the tabulated annihilation spectra by evaluating these at half the relevant particle mass. Furthermore, for kinematic reasons, if we produce two standard model particles, each of rest mass m_A , then for decay we enforce the DM particle mass to obey $m_\chi > 2m_A$, whereas for annihilation this is $m_\chi > m_A$.

4.3 Results

In this section we describe the constraints for all channels considered – s -, p - and d -wave. First we present the results on the s -wave constraints, describing the whole statistical analysis pipeline step by step, given that the posterior analysis for this case is more involved than for p - or d -wave case due to the presence of the D factor template for which we actually detect a non-zero value at 2σ significance level as discussed below.

4.3.1 The s -wave constraints

In Fig. 4.4, we show the corner plot for the first stage of our inference, where we infer $A_i^{J^{(\ell=0)}}$ and A_i^D simultaneously. We emphasise that we fit a different $A_i^{J^{(\ell=0)}}$ and A_i^D to each energy bin, i , and CSiBORG simulation. In this example, we consider simulation number 7444 (as given in [186]) and the energy range 30–50 GeV (energy bin 9). We see that the parameters corresponding to the isotropic, galactic and point-source contributions are all approximately unity, as one would expect. For this energy bin and CSiBORG simulation we see that there is no evidence for a contribution to the gamma-ray flux proportional to either the $J^{(\ell=0)}$ or the D factor. We note that $A_i^{J^{(\ell=0)}}$ and A_i^D are highly degenerate, such that a large value of $A_i^{J^{(\ell=0)}}$ corresponds to a small A_i^D . For our fiducial analysis, we therefore choose to set one of these parameters equal to zero at a time; i.e. the inference to place constraints on $\langle\sigma v\rangle$ will assume $\forall i A_i^D = 0$ and for Γ we assume $\forall i A_i^{J^{(\ell=0)}} = 0$. This will make our constraints conservative (see Section VI A.1 in [132]).

⁶https://fermi.gsfc.nasa.gov/ssc/data/analysis/scitools/source_models.html

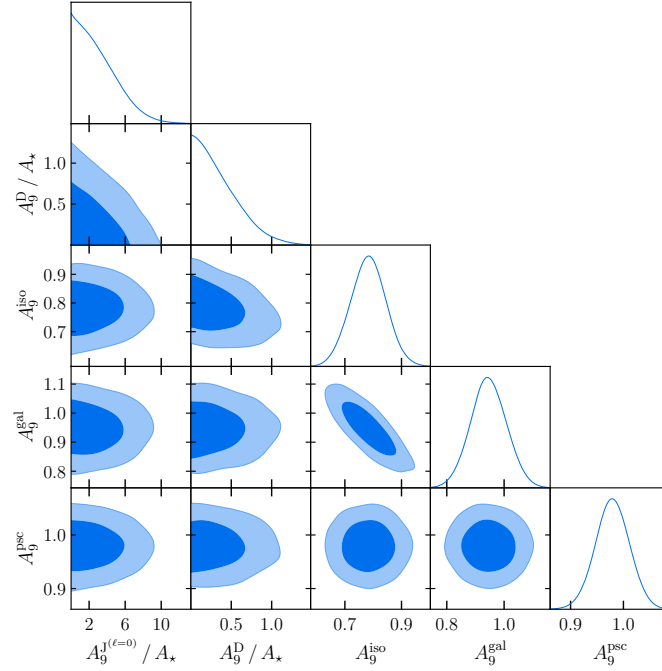


Figure 4.4: Posterior distributions for CSiBORG simulation 9844 of the parameters describing the gamma-ray flux in the energy range 30 – 50 GeV for the case of s -wave ($\ell = 0$). We include templates proportional to the $J^{\ell=0}$ factor ($A_9^{J^{\ell=0}}$) and D factor (A_9^D), as well as isotropic (A_9^{iso}), galactic diffuse (A_9^{gal}), and point source (A_9^{psc}) contributions, and define $A_\star \equiv 10^{-22} \text{ cm}^{-2} \text{ s}^{-1} \text{ MeV}^{-1}$. The contours show the 1 and 2σ confidence intervals.

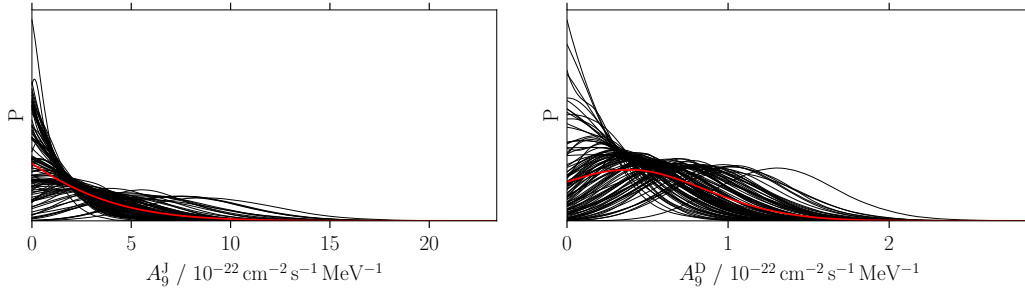


Figure 4.5: The corresponding one-dimensional posterior distributions, complementing the posterior from Fig. 4.4, on the coefficients describing the flux proportional to the (left panel) J factor and (right panel) D factor in the energy range 30 – 50 GeV. Each black line gives the posterior distribution for an individual CSiBORG simulation, and the red line is the mean of these, i.e., the final posterior distribution.

We note that A_i^{iso} is strongly degenerate with A_i^{gal} , which is to be expected since both describe large-scale features across the sky. If we used exactly the same selection criteria as the *Fermi* analysis which produced the non-DM templates, then A_i^{gal} , A_i^{iso} and A_i^{psc} would all have a mean of unity. This is not true here because the isotropic template is calibrated for latitudes $10^\circ < |\lambda| < 60^\circ$, whereas we fit our template to $|\lambda| > 30^\circ$. In general, we find A_i^{iso} to be slightly smaller than 1. This is more prominent in the higher energy bins; we find that A_i^{iso} is closer to unity at lower energy. We verify that this is not due to the addition of the $J^{(\ell=0)}$ or D factor templates by rerunning the analysis with $A_i^{\text{D}} = A_i^{J^{(\ell=0)}} = 0$ and find that A_i^{iso} remains less than one.

We generate such MCMC chains for each of the 101 CSiBORG simulations, and we plot the resulting one-dimensional posterior distributions for $A_i^{J^{(\ell=0)}}$ and A_i^{D} in Fig. 4.5. Knowing that each CSiBORG simulation is a fair Monte Carlo sample, the final posterior distribution on A_g^{X} is simply the average of each individual probability distribution, which yields the red lines in the figure. When marginalised over the BORG chain, we again find $A_i^{J^{(\ell=0)}}$ and A_i^{D} are consistent with zero for this energy bin.

The posterior of $A_i^{J^{(\ell=0)}}$ and A_i^{D} is shown in Fig. 4.6. The posterior is over all considered energy bins, marginalised over all other contributions to the gamma-ray sky and over the uncertainties involved in producing maps of the $J^{(\ell=0)}$ and D factors. We indicate the maximum posterior points by the circles and 1σ confidence intervals by the error bars. For a given DM mass and channel, these posteriors can be trivially transformed into constraints on $\langle\sigma v\rangle$ or Γ for a given energy bin. We then simply multiply the posteriors from each bin to determine our final constraint on these parameters.

In Fig. 4.7, we plot the 95% upper limit on $\langle\sigma v\rangle$ as a function of DM particle mass, m_χ , for an annihilation which solely produces particle-antiparticle pairs of a single type, but for any standard-model quark, charged lepton or gauge boson (except photons). We compare these constraints to the thermal relic cross section ($\langle\sigma v\rangle_{\text{th}} \approx 3 \times 10^{-26} \text{ cm}^3 \text{ s}^{-1}$), such that, if the curve falls below this value in Fig. 4.7, then we rule out DM being a thermal relic for the corresponding mass and annihilation channel at 95% confidence. For all annihilations producing gluons or quarks less massive than the bottom quark, we see that, if DM is a thermal relic, it should be more massive than $\sim 7 \text{ GeV}/c^2$ since we rule out smaller masses. We are unable to rule out the thermal relic cross section at any mass for production of bottom quarks, top quarks, W bosons or Z bosons. Our constraints for lepton production are much weaker at a given particle mass, such that our constraints for electron or muon production do not cross $\langle\sigma v\rangle_{\text{th}}$. We rule out τ production for $m_\chi \lesssim 6 \text{ GeV}/c^2$ at this cross section.

Turning our attention to DM decay, Fig. 4.8 shows the inferred decay rate, Γ , for different decay channels as a function of m_χ . Contrary to our analysis of DM annihilation, we find that for the majority of channels we infer a nonzero Γ at over 2σ confidence for at least some m_χ (corresponding to nonzero A_i^{D} in Fig. 4.6). The results are relatively insensitive to the DM particle mass, provided m_χ is above the threshold for production.

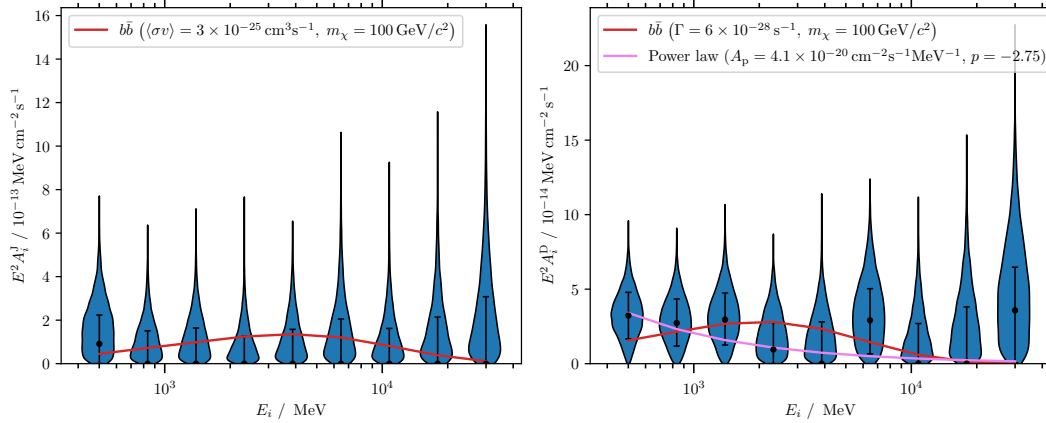


Figure 4.6: The resulting s -wave channel one-dimensional posterior distributions on the coefficients describing the flux in each energy bin, i , proportional to (left) the J factor and (right) the D factor. The black points correspond to the maximum likelihood points and the error bars show the 1σ confidence interval. For reference, we plot the expected A_i^J and A_i^D for DM annihilation and decay, respectively, via the $b\bar{b}$ channel for a particle of mass $m_\chi = 100 \text{ GeV}/c^2$ with a thermally averaged cross section of $\langle\sigma v\rangle = 3 \times 10^{-25} \text{ cm}^3 \text{ s}^{-1}$ and decay rate $\Gamma = 6 \times 10^{-28} \text{ s}^{-1}$. We also plot A_i^D if the spectrum was due to a power law of amplitude $A_p = 4.1 \times 10^{-20} \text{ cm}^{-2} \text{ s}^{-1} \text{ MeV}^{-1}$ and index $p = -2.75$.

For the $b\bar{b}$ channel, we find the inferred Γ is $\sim 6 \times 10^{-28} \text{ s}^{-1}$, which corresponds to approximately one decay per Hubble time in a volume $\sim 280 \text{ km}^3$ at mean cosmological density if $m_\chi = 100 \text{ GeV}/c^2$. This is around the smallest Γ that has been constrained by any previous study (see Section 4.5.6). The inferred Γ for decay to the lightest charged leptons is approximately an order of magnitude larger than this.

To determine the overall detection significance, we compute the coefficient for the total flux across all energy bins which multiplies the D factor

$$A_{\text{tot}}^D \equiv \sum_i A_i^D \Delta E_i = 1.02_{-0.28}^{+0.24} \times 10^{-16} \text{ cm}^2 \text{ s}^{-1}. \quad (4.67)$$

Simply dividing the best fit value by the lower error would suggest that our detection of a contribution to the gamma-ray sky proportional to the D factor has a significance of 3.6σ when averaged over all available energies. Since our posterior is non-Gaussian, we wish to compute this significance through other methods. We compute the maximum likelihood for each A_i^D and, since each energy bin is treated as independent, the maximum likelihood for A_{tot}^D is the product of these values. We compare this to the likelihood for $A_i^D = 0$ and find the change in log likelihood between these two cases is $\Delta\chi^2 \equiv 2\Delta \ln \hat{\mathcal{L}} = 11.2$, which is equivalent to 3.3σ for a Gaussian likelihood or a change in the Bayesian Information Criterion (BIC) [230] of 9.0, if one takes the $A_{\text{tot}}^D \neq 0$ model as having one

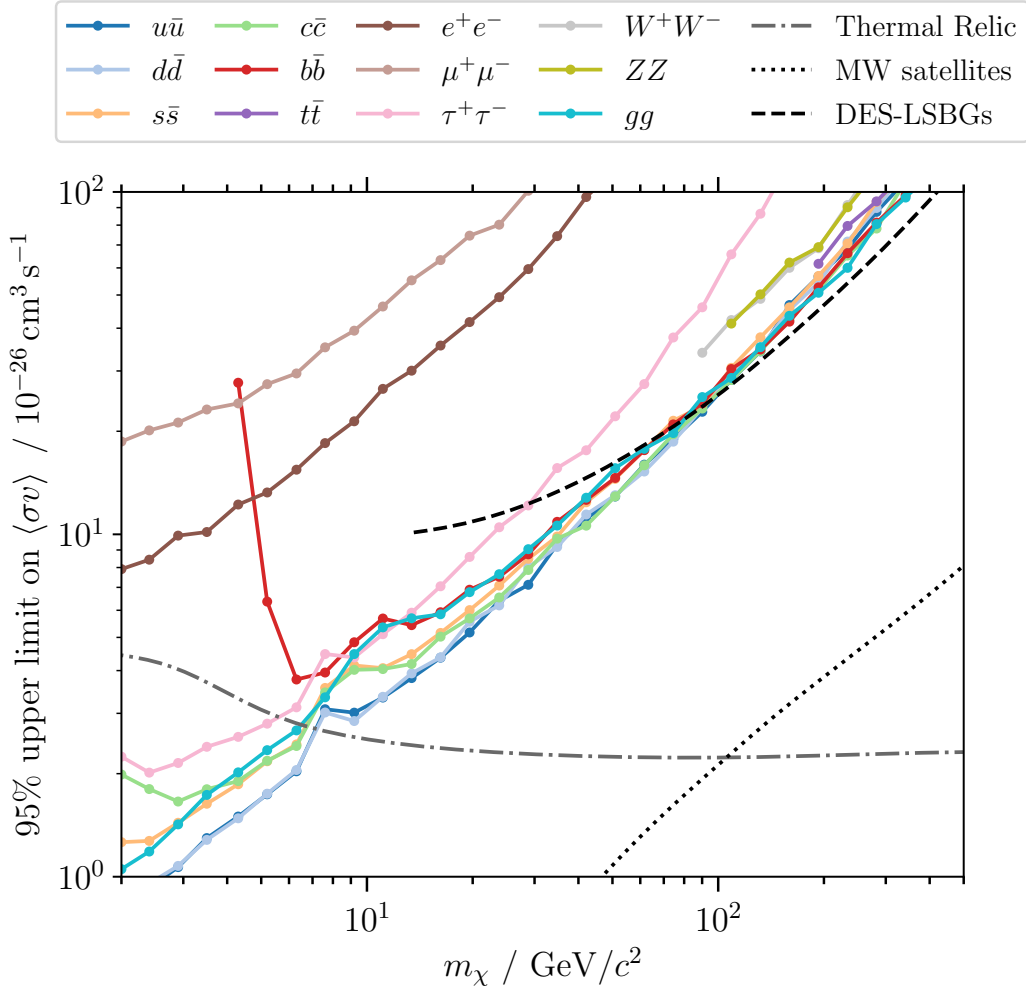


Figure 4.7: Constraints on DM annihilation cross section, $\langle\sigma v\rangle$, as a function of particle mass, m_χ , for different annihilation channels. The dot-dashed grey line is the expectation for a thermal relic, $\langle\sigma v\rangle_{\text{th}}$, as calculated by [60]. All points below this line rule out the thermal relic cross section at 95% confidence for the corresponding mass and channel. The dotted black line is the constraint obtained by [169] from Milky Way satellites for the $b\bar{b}$ channel; we see our constraints are approximately an order of magnitude less stringent. The dashed black line shows the constraints for the $b\bar{b}$ channel derived from the cross-correlation between *Fermi*-LAT and the Dark Energy Survey Y3 low surface brightness galaxy sample (DES-LSBGs) [229]. Our field-level inference improves the constraints from large-scale structure by approximately a factor of 2 at $m_\chi = 10 \text{ GeV}/c^2$.

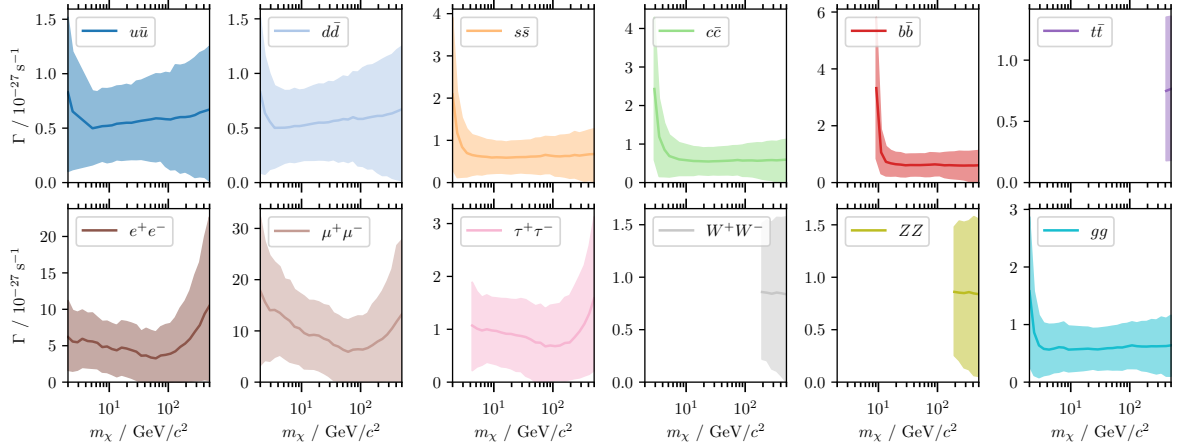


Figure 4.8: Constraints on DM decay rate, Γ , as a function of particle mass, m_χ , for different decay channels. The solid lines are the median values and the bands show the 95% confidence regions. We do not infer m_χ , which means that every constraint is conditioned on the corresponding particle mass. For some decay channels and some masses, we see that Γ is inferred to be nonzero; however (as shown in Section 4.4.2), we find that a power-law model better describes the spectrum, suggesting that this flux is not in fact due to DM decay. We note that for some channels we cannot probe the full mass range due to the requirement that m_χ is at least as large as the sum of the masses of the decay products.

more parameter. In Section 4.4.2 we ask whether this is due to DM decay, finding that a non-DM spectrum is preferred by the data.

4.3.2 The p -wave and d -wave constraints

As in previous section, the first step of our inference is to infer the amplitudes of the $J^{(\ell>0)}$ factor templates, A_i^J , for each energy bin, i . In Fig. 4.9 we plot the one-dimensional posterior distributions of these parameters and see that these coefficients are consistent with zero at 1σ confidence for all energy bins for both p - ($\ell = 1$) and d -wave ($\ell = 2$) annihilation. In the second step we fit this inferred spectrum to a dark matter annihilation model and hence can constrain a_ℓ . For reference, in Fig. 4.9 we plot example spectra for annihilation via the $b\bar{b}$ channel for $m_\chi = 100 \text{ GeV}/c^2$ at a fiducial a_ℓ .

In Fig. 4.10 we convert these posteriors to 95% upper limits on the p - and d -wave self-annihilation cross-section ($\sigma v = a_\ell (v/c)^{2\ell}$) as a function of dark matter particle mass, m_χ , for the production of any standard-model quark, charged lepton or gauge boson (except photons). We do not attempt to constrain m_χ , and instead find constraints on

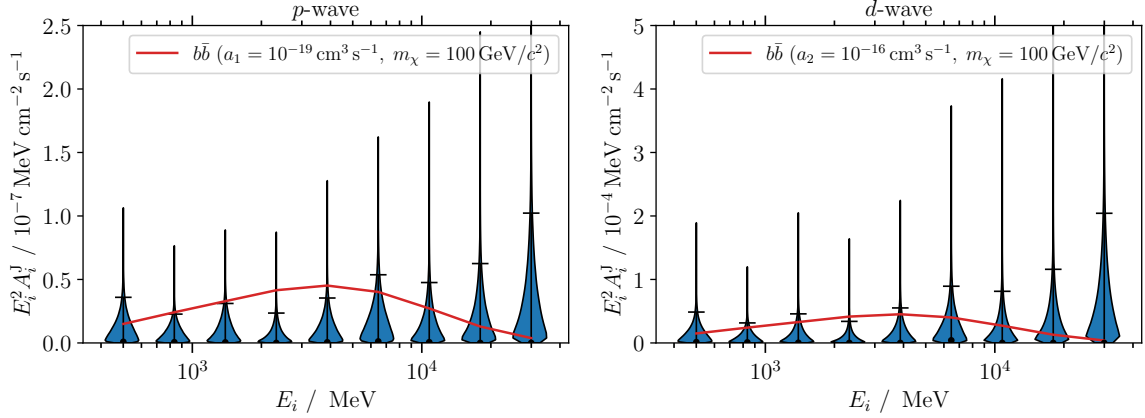


Figure 4.9: Posterior distributions of the amplitudes of the J factor templates for each energy bin, i , for p -wave (left) and d -wave (right) annihilation. The black points are the maximum posterior points and the error bars indicate the 1σ confidence intervals. For reference we also plot the expected A_i^J for dark matter annihilation via the $b\bar{b}$ channel for dark matter particles of mass $m_\chi = 100 \text{ GeV}/c^2$, with $a_1 = 10^{-19} \text{ cm}^3 \text{ s}^{-1}$ (p -wave) and $a_2 = 10^{-16} \text{ cm}^3 \text{ s}^{-1}$ (d -wave). All fluxes are consistent with zero, indicating no detection of dark matter p -wave or d -wave annihilation.

a_ℓ conditioned on this variable for the range $m_\chi \in [2, 500] \text{ GeV}/c^2$. In all cases we find no evidence for p -wave or d -wave dark matter annihilation, as all the fluxes in Fig. 4.9 are consistent with zero. For dark matter annihilating to $b\bar{b}$, if $m_\chi = 10 \text{ GeV}/c^2$, we constrain $a_1 < 2.4 \times 10^{-21} \text{ cm}^3 \text{ s}^{-1}$ and $a_2 < 3.0 \times 10^{-18} \text{ cm}^3 \text{ s}^{-1}$ at 95% confidence. For an annihilation which forms lighter quarks, we obtain practically identical constraints, whereas lepton production yields constraints which are approximately an order of magnitude less stringent. For large $m_\chi \gtrsim 100 \text{ GeV}/c^2$, our upper limits for the production of top quarks or W or Z bosons are similar to those for lighter quarks, but these processes cannot occur at lower m_χ so we do not obtain constraints for smaller dark matter particle masses.

Note that, in Fig. 4.10, the constraints weaken as m_χ approaches m_A , if the annihilation product is $A\bar{A}$. Near this value, the process is strongly suppressed and hence the constraints no longer lie within the range of the plot, so we do not plot these points.

For comparison, for p -wave, in Fig. 4.10 we also plot the constraints one obtains for the $b\bar{b}$ channel from dwarf spheroidals. As anticipated from the much smaller velocity dispersion in dwarf galaxies, our constraints are significantly tighter, by approximately two orders of magnitude for p -wave annihilation, and seven orders of magnitude for d -wave. This demonstrates the increased importance of large halo masses for velocity-

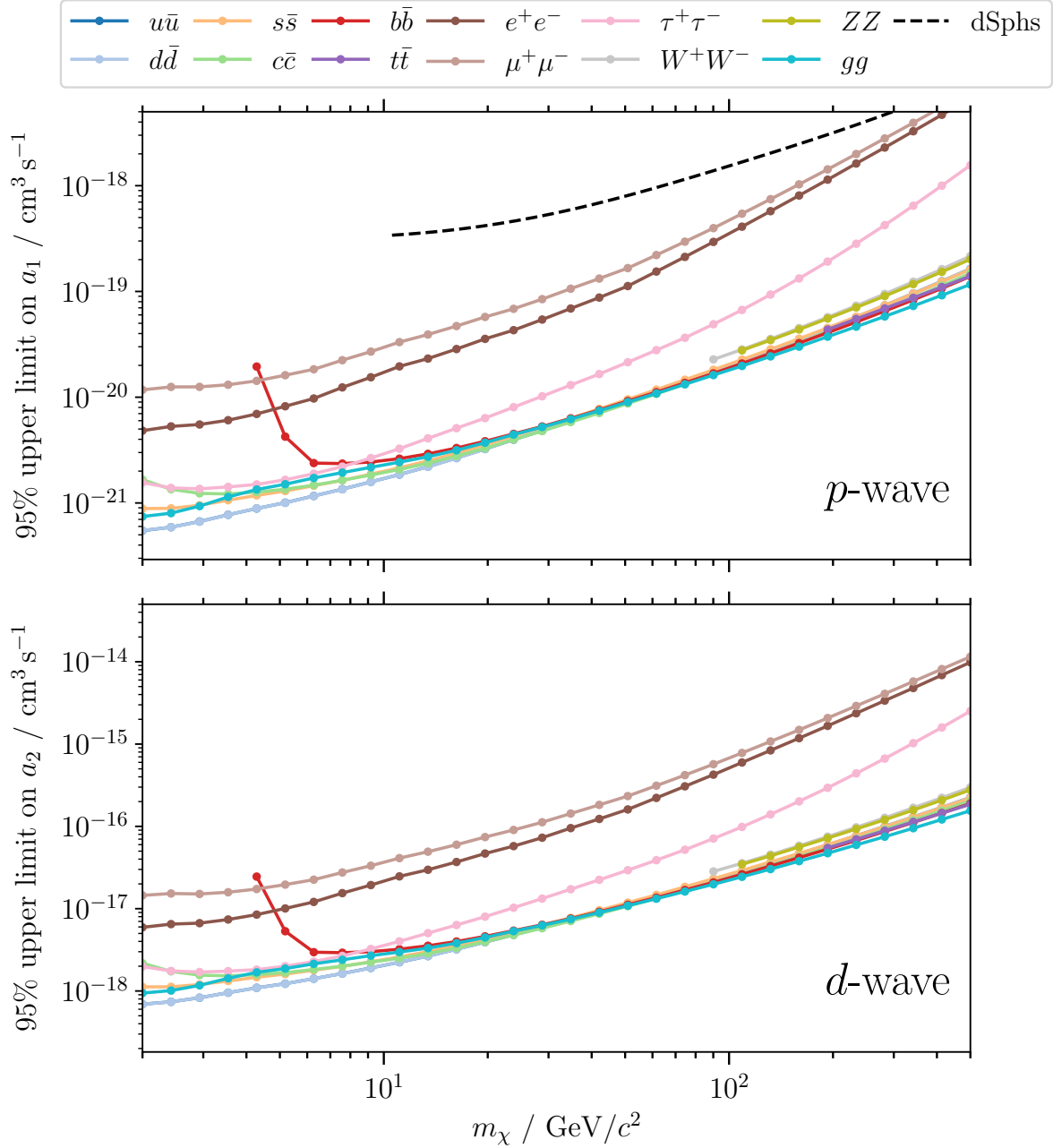


Figure 4.10: 95% upper limit on the dark matter p -wave (upper, $\ell = 1$) and d -wave (lower, $\ell = 2$) annihilation cross-section parameter, a_ℓ , where the annihilation cross-section is $\sigma v = a_\ell(v/c)^{2\ell}$. The solid lines and points are from this work. Each constraint is derived at a fixed dark matter particle mass m_χ as indicated on the x -axis. The red dotted line is a constraint obtained from dwarf spheroidal (dSph) satellites of the Milky Way [231, 232] for the $b\bar{b}$ channel. We find that our constraints are approximately two orders of magnitude tighter than those from dSphs due to the larger velocity dispersions in massive extragalactic halos. The dSph constraint for d -wave is off the top of the plot. The brown dashed line is the constraint for the e^+e^- channel, directly detecting the resulting electrons and positrons, assuming they are produced by the galactic halo [233]. This is slightly tighter than our e^+e^- constraint.

dependent annihilation, whereas for s -wave annihilation the close proximity of these objects more than compensates for their small masses, leading to tighter constraints (see Fig. 4.7). We perform a more detailed comparison to the literature in Section 4.5.6.

4.4 Interpretation of the results

In this section we investigate which objects and observations drive our results and whether there are non-DM explanations for the signal proportional to the D factor.

4.4.1 Annihilation

s -wave channel

For each channel and m_χ , we compute the change in log likelihood between $\langle\sigma v\rangle = 0$ and the 1σ constraint on $\langle\sigma v\rangle$ separately for each energy bin to determine which energy range is dominating our constraint. For DM particle masses $m_\chi \lesssim 10 \text{ GeV}/c^2$, our constraints are driven by the first three energy bins. For these masses, we expect there to be very few photons produced at high energies at fixed $\langle\sigma v\rangle$, so these bins are unable to constrain $\langle\sigma v\rangle$ as very large values are required to produce an appreciable flux. As we move to higher masses, we notice the effect of the data at higher energy, such that at the highest masses we find that the sixth energy bin (6.5 – 10.8 GeV) is the most constraining. We find a similar trend if we compare using the 95% confidence limit instead of the 1σ constraint.

To determine which halos drive our constraints, we produce several J -factor maps where each one is only due to objects in a given mass range. We create separate maps for halos in moving bins of width $\Delta \log_{10}(M_h/M_\odot) = 1$. We rerun the inference for a single representative CSiBORG simulation (9844) separately for each of these mass bins; i.e., we assume that only a single mass bin contributes to the total J factor. We plot the constraint on $\langle\sigma v\rangle$ as a function of halo mass in Fig. 4.11 and observe that the tightest constraints are obtained for halos in the range $\sim 10^{13.5} - 10^{14.5} M_\odot$. If one studied a single object at a fixed distance, then the most massive halo would give the tightest constraints since it has the largest J factor. However, such massive objects are rare, so there is a compromise between having many objects of a given mass across the sky and those objects having a large J factor. Given the tight constraints one can obtain with dwarf galaxies in the Local Group, it is perhaps not surprising that the inclusion of lower mass objects can lead to an improvement in the upper limit on $\langle\sigma v\rangle$. The inclusion of these structures in this work was possible due to the use of constrained simulations, which provide plausible realisations of these halos given the ICs that are constrained on large scales. As a result, Fig. 4.11 shows the types of objects in the nearby universe one should target to extract maximum information about DM annihilation.

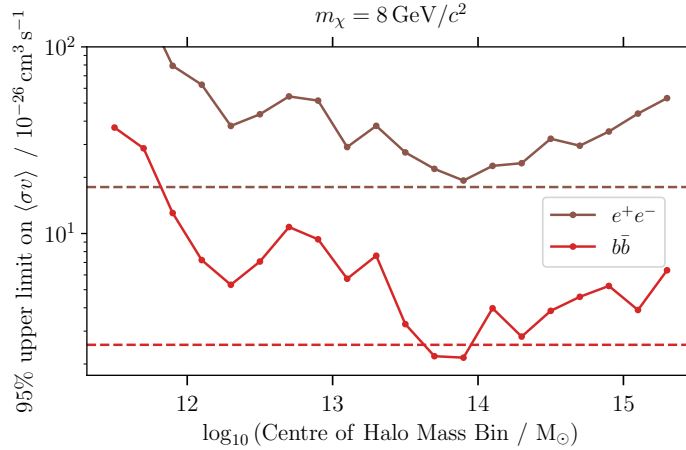


Figure 4.11: Constraints on the s -wave self-annihilation cross section, $\langle\sigma v\rangle$, to either bottom quarks or electrons from halos within a given mass range, where we consider bins in halo mass of width $\Delta \log_{10}(M_h/M_{\odot}) = 1$. Here we use only a single CSiBORG simulation (9844); the 95% upper limits on $\langle\sigma v\rangle$ for this simulation if we use all halos are indicated by the dashed horizontal lines. Our constraints are dominated by halos of mass $\sim 10^{13.5} - 10^{14.5} M_{\odot}$.

Since we found a nonzero flux proportional to the D factor, we rerun the analysis for all CSiBORG simulations but infer both A_i^D and $A_i^{J^{(\ell=0)}}$ simultaneously. In this way, when making our constraint on $\langle\sigma v\rangle$, we now marginalise over this contribution. Note that in this marginalisation we do not assume a spectral form for the D factor template, such that we marginalise over any source whose spatial distribution is proportional to the local DM density, which may or may not be DM decay. As anticipated in Section 4.3, we find that our constraints become tighter, such that our upper limit on $\langle\sigma v\rangle$ is typically over a factor of 2 smaller. This is as expected: the negative degeneracy between $A_i^{J^{(\ell=0)}}$ and A_i^D (Fig. 4.4) means that, if we allow $A_i^D > 0$, we must reduce $A_i^{J^{(\ell=0)}}$ so that the total flux from these two contributions is approximately constant. When marginalising over A_i^D , this will result in a posterior on $A_i^{J^{(\ell=0)}}$ which is necessarily narrower. Since these results are tighter than when we set $A_i^D = 0$, we choose to report the latter as our fiducial results so that our conclusions are conservative.

p - and d -wave channels

To determine which halos are responsible for our constraints, we produce J factor maps for halos only within a given mass range and run the inference using only those halos. For computational convenience, we perform this on one representative CSiBORG simulation only (9844). The results of this inference are plotted in Fig. 4.12, where we use moving

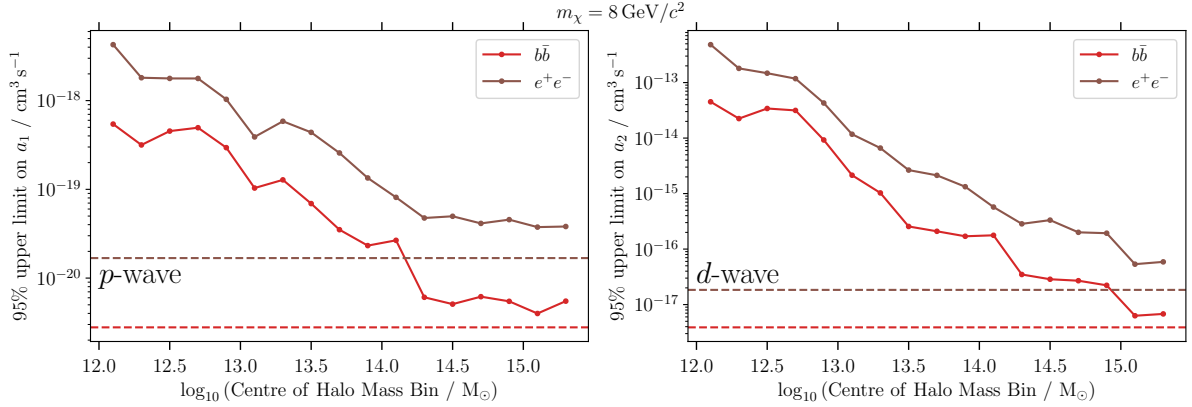


Figure 4.12: Constraints on p -wave (left, $\ell = 1$) and d -wave (right, $\ell = 2$) annihilation cross-section, $\sigma v = a_\ell (v/c)^{2\ell}$, from halos in a given mass range from a single CSiBORG simulation (9844). Each mass bin has a width $\Delta \log_{10} (M_h/M_\odot) = 1$, and we use a sliding window. We plot the 95% upper limits on a_ℓ for the $b\bar{b}$ and e^+e^- channels; the dashed horizontal lines are the constraints obtained from all halos together. We see that our constraints are driven by the most massive halos and that this is more important for d -wave annihilation due to the stronger dependence of annihilation rate on the relative velocity between dark matter particles. For comparison to s -wave case see Fig. 4.11.

mass bins of width $\Delta \log_{10} (M_h/M_\odot) = 1$ and consider dark matter particles of mass $m_\chi = 8 \text{ GeV}/c^2$ annihilating via the $b\bar{b}$ or e^+e^- channels. We see that our constraints are dominated by the most massive objects in our simulations, i.e. halos of mass $M_h \sim 10^{14-16} M_\odot$. This is due to both the larger densities and velocity dispersions in massive objects, as was predicted in [192] (see their Fig. 3). If we continued Fig. 4.12 to larger M_h we would see that the constraints would begin to weaken, due to the lack of very massive objects in our simulation. We observe that the constraining power as a function of mass for p -wave annihilation plateaus at an earlier halo mass than for d -wave annihilation due to the weaker dependence on the relative velocity between dark matter particles, which increases with mass. Qualitatively similar results are seen for other channels and dark matter particle masses. Again, this is consistent with the expectations of [192]. We find that the total J factors we obtained, for the 128 most massive halos of one of our CSiBORG realisation (9844), agree to within factor of ~ 5 when compared to the expression from Eq. (3.3) in [192]. This is a consequence of more detailed modelling of the J -factor in this work; for example, if the authors of [192] chose a different mass definition, given the strong dependence of the p - and d -wave annihilation channels on total mass of the object, this difference could be alleviated.

To assess which energy bins drive our constraints, for each channel and m_χ we calcu-

late the log-likelihood at the 95% upper limit on a_ℓ for each energy bin and compare to the values for $a_\ell = 0$. For both p - and d -wave annihilation, we find that the first three energy bins are the most constraining for $m_\chi \lesssim 30 \text{ GeV}/c^2$, whereas the fourth energy bin becomes particularly important for higher masses. This is the same behaviour as for s -wave annihilation, which is attributed to the lower expected photon counts in the higher energy bins, and therefore weaker constraining power. As one increases m_χ , one expects more photons at higher energy, and thus the relative importance of these bins increases.

4.4.2 Decay

In Fig. 4.8 we found a nonzero DM decay rate is compatible with the observed gamma-ray sky at over 2σ confidence for a range of DM masses and decay channels. More conservatively, one would say that we find a signal which is proportional to the D factor; i.e., the emitted flux from some region appears to be proportional to the local density, and is compatible with the spectrum of DM decay. This source does not necessarily need to be DM decay, which can be investigated by fitting the inferred spectrum to a different model. For this we choose a power-law profile, such that the parameter A_i^D arises from integrating the spectrum

$$\frac{dN}{dE_\gamma} = A_p \left(\frac{E_\gamma}{E_0} \right)^p, \quad (4.68)$$

across energy bin i , where we normalise to $E_0 \equiv 1 \text{ GeV}$. We place broad, uniform priors on A_p and p in the range $[0, 10^{-18}] \text{ cm}^{-2}\text{s}^{-1}\text{MeV}^{-1}$ and $[-5, 2]$, respectively, and find

$$\begin{aligned} A_p &= (4.1 \pm 1.5) \times 10^{-20} \text{ cm}^{-2}\text{s}^{-1}\text{MeV}^{-1}, \\ p &= -2.75_{-0.46}^{+0.71}, \end{aligned} \quad (4.69)$$

where the limits are at 1σ confidence. To enable a comparison, we plot the spectrum for this model and a DM decay model in Fig. 4.6, where we see that the power law fits better at most energies.

To assess the relative goodness of the fit of the two models we compute the BIC. Since we have set deliberately wide priors on our model parameters, ratios of the Bayesian evidence are difficult to interpret. For all channels and masses, we find that the BIC prefers the power-law spectrum, with $\text{BIC} \geq 1.3$ (the bound is saturated for the gluon channel). Decays via the $t\bar{t}$ channel are least preferred by the data, with $\text{BIC} \geq 4.0$. We therefore conclude that, although we do find an excess of gamma-ray flux that traces the density of DM, its spectrum is fit marginally better by a power law, so this is not evidence for DM decay.

This conclusion is consistent with previous works studying the origin of the residual gamma-ray flux. The cross-correlation of the gamma-ray sky with galaxy catalogues has been detected at $2 - 4\sigma$ [234], and the spectral index of the 1-halo contribution was

found to be -2.7 if a single power law is assumed, which is consistent with our inferred index and has a similar significance of detection. Moreover, analysis of the angular power spectrum of the gamma-ray sky [235] suggests a component that can be modelled as a double power law with an exponential cutoff, with power-law indices -2.55 ± 0.23 and -1.86 ± 0.15 , the former of which is consistent with our result. [235] note that this spectrum is compatible with blazarlike sources being the dominant component at these energies. Since we detect a nonzero contribution with the same spatial variation as our D factor maps, our work suggests that these excesses could be due to sources with a linear bias with respect to the local DM density.

4.5 Systematic uncertainties

In this section we describe the systematic uncertainties associated with our analysis and quantify them for all of the considered s -, p - and d -wave channels.

4.5.1 J and D factor calculation

s -wave channel

In Section 4.2.2 we computed the D factor and nonhalo contribution to the J factor by smoothing simulation particles onto a grid with a kernel inspired by SPH. Besides this kernel, we also consider the CIC density assignment in order to quantify the impact of the kernel choice on our constraints. The median change in the constraint on $\langle\sigma v\rangle$ is 2% for the $b\bar{b}$ channel if we change to this kernel. For the DM decay inference, we find the median and 95% upper limit on Γ change by a median of 16% and 15%, respectively, for this channel. The inferred value of p changes only by 0.01 when we change to the CIC kernel. If the low density regions were driving our constraints, then one would expect large differences between the two procedures, since these regions have the fewest simulation particles and the two kernels have different noise properties for low particle counts. However, we do not see this since the expected flux is highest in the high density regions and the low density regions are relatively unconstraining.

For computational convenience, for our fiducial analysis we chose a HEALPIX resolution of `nside=256`. We rerun the analysis at coarser resolution (`nside=128`) and find that our constraints on $\langle\sigma v\rangle$ weaken by a median change of 48% across all masses for the $b\bar{b}$ channel. The median and 95% upper limit on Γ change by a median of 13% and 10% respectively for the $b\bar{b}$ channel, and the inferred value of p changes by only 0.03 when we lower the HEALPIX resolution. It is unsurprising that the D factor analysis is less affected by this choice; for the J factor our template is dominated by high density peaks in the DM density field, since the flux is proportional to the square of the density. By using a higher resolution map, one can localise these peaks better to obtain tighter

constraints if these are not aligned with peaks in the observed gamma-ray sky.

The J and D factor maps were calculated for each of the 101 CSiBORG simulations. By utilising the full suite, we marginalise over the uncertainties in the constrained density modes from both the BORG algorithm and the unconstrained, small-scale modes. To verify that we have a sufficiently large number of simulations to achieve this, we rerun our analysis 100 times for the $b\bar{b}$ channel using 50 randomly selected simulations to determine a bootstrap uncertainty on our constraints. The standard deviation of the 95% upper limit on $\langle\sigma v\rangle$ has a median value of 19% when considering all masses. The inferred Γ has a median bootstrap uncertainty of 4%. The uncertainty on the inferred power-law index, p , is 0.05 and the fractional bootstrap uncertainty on A_p is 4%, which are small compared to the uncertainties we quote in Eq. (4.69). We therefore conclude that the number of constrained simulations is adequate.

Since the gamma-ray flux from DM annihilation is proportional to the square of the density, the substructure of DM halos is an important contribution that one must consider; if one computes the angular power spectrum for the J factor, C_ℓ , one finds that $(2\ell + 1)C_\ell$ is approximately constant at the smallest scales considered in this work. Usually this substructure is modelled as a mass-dependent multiplicative boost factor, e.g. [236] and uncertainties captured by looking at the extreme values of the boost for given masses [179]. We, on the other hand, capture substructure and its uncertainty through CLUMPY's probabilistic approach to substructure modelling. This led us to use a non-Poisson likelihood, since we introduced uncertainties, σ_{jp} , on the Poisson means. To evaluate the impact of this choice, we rerun the analysis assuming a Poisson likelihood by setting $\sigma_{jp} = 0$. We find that our constraints typically change by a few percent across all channels and masses, indicating that the impact of this uncertainty is negligible. However, we note that this could not have been known *a priori*. Although the fractional uncertainties are small near the centres of halos, this is not true in the outskirts, motivating our thorough treatment of uncertainties.

For simplicity, we previously neglected the uncertainty that arises due to stochasticity in the mass-concentration relation. We find our constraints are not very sensitive to the scatter in this relation. For the $b\bar{b}$ channel across all masses, the median change in the constraint on $\langle\sigma v\rangle$ is 1.4% if this uncertainty is included.

Another source of systematics due to substructure modelling might be driven by the breakdown of our assumption that the uncertainty in the J factor is Gaussian. Namely, as the considered mass of the clump grows, the total number of such clumps within the host halo decreases. Therefore, it is expected that at some point we transition from the Gaussian into a Poisson regime [237]. Furthermore, it is not obvious that the contribution to the J factor from these more massive clumps will not outshine the cumulative contribution of the lower-mass clumps. To check for this, we use the `-h5` module of the CLUMPY package to explicitly draw substructure realisations for a typical

halo ($M_h \approx 5 \times 10^{13} M_\odot$) from the CSiBORG simulations. We modify Eq. (4.39) such that

$$\langle J_p \rangle = \langle J_{\text{cont},p}(M_{\text{th}}) \rangle + \langle J_{\text{drawn},p}(M_{\text{th}}) \rangle, \quad (4.70)$$

and

$$\begin{aligned} \langle J_{\text{cont},p}(M_{\text{th}}) \rangle &= J_{\text{sm},p} + \langle J_{\text{subs},p}(M_{\text{th}}) \rangle \\ &+ \langle J_{\text{cross},p}(M_{\text{th}}) \rangle, \end{aligned} \quad (4.71)$$

where we introduce

$$\langle J_{\text{drawn},p}(M_{\text{th}}) \rangle = \frac{1}{N_{\text{ds}}} \sum_i^{N_{\text{ds}}} J_{\text{drawn},pi}(M_{\text{th}}), \quad (4.72)$$

with N_{ds} being the total number of explicit realisations of the clumps with a mass above a given mass threshold, M_{th} . The quantities $\langle J_{\text{subs},p}(M_{\text{th}}) \rangle$ and $\langle J_{\text{cross},p}(M_{\text{th}}) \rangle$ from Eq. (4.71) are obtained by replacing the upper limit of the clump mass distribution by M_{th} , i.e., replacing M_2 with M_{th} in Eq. (4.31). To estimate $\langle J_{\text{drawn},p}(M_{\text{th}}) \rangle$, we run $N_{\text{ds}} \approx 1000$ explicit realisations of substructure clumps for a typical halo, requiring that we capture fluctuations in the value of $J_{\text{sm},p}$ – the leading contribution to the total J factor of the halo – at the percent level. In other words, any clump whose contribution to the given pixel will induce a fluctuation to the value of $J_{\text{sm},p}$ of the order of $\sim 1\%$ will be explicitly drawn onto the HEALPIX grid. This is equivalent to taking one sample from Eq. (4.25), but with a modified mass range of the mass function, and setting the lower limit for this draw to be M_{th} . For more details see section 2.4.3 of [199].

For this experiment, we selected a HEALPIX resolution of $\text{nside} = 1024$, corresponding to the *Fermi*-LAT angular resolution. The corresponding threshold mass for this setup translates to $M_{\text{th}} = 5.3 \times 10^9 M_\odot$ for our chosen halo. We find that

$$\langle J_{\text{drawn},p}(M_{\text{th}}) \rangle \sim 0.04 \langle J_{\text{cont},p}(M_{\text{th}}) \rangle, \quad (4.73)$$

which justifies our starting assumption of treating the substructure contribution to the total J factor as Gaussian, since the “drawn” (Poisson) component is subdominant compared to the “continuous” (Gaussian) contribution.

Note, however, that choosing a smaller M_{th} , i.e., looking at even smaller fluctuations of $J_{\text{sm},p}$, would lead to probing the even lower-mass end of the substructure mass function, which would, of course, alter the ratio of the “drawn” and “continuous” components. However, going below this limit would already enter into a regime where drawing $10^4 - 10^5$ clumps from the corresponding version of Eq. (4.25) will be necessary, which is computationally expensive and well within the validity of the Gaussian approximation. As a comparison, using $M_{\text{th}} \approx 5.3 \times 10^9 M_\odot$ required around $\sim 10^3$ draws. Throughout this section we assumed the same DM profile parameterisation and mass-concentration relation as in the fiducial inference. The conclusions are unchanged for the same halo using the Einasto profile.

p- and *d*-wave channels

In Equation (4.62) we marginalised over uncertainties in the large-scale density modes which are constrained from the BORG algorithm, as well as smaller-scale unconstrained modes which were added as white noise in the initial conditions. In our fiducial analysis we use the full suite of 101 CSiBORG simulations, however to verify this is a sufficiently large number, we rerun our analysis one hundred times using fifty randomly selected CSiBORG simulations (with repeats) each time. This yields an uncertainty on the 95% upper limit of a_ℓ of no more than 5.9% and 6.5% for *p*-wave and *d*-wave annihilation, respectively, when considering all channels and particle masses separately. The median change in the constraint across these runs at a given mass and channel changes by no more than 1.2% compared to the fiducial analysis. Thus, we conclude that this is an appropriate number of simulations.

For our *p*-wave analysis, we assumed circular orbits for dark matter particles ($\beta = 0$), leading to the radial velocity dispersion given in Eq. (4.45). Similar expressions exist for alternative values of β , thus to determine the impact of this assumption on our constraints, we compute J factor maps assuming $\beta = 0.5$ or $\beta = 1$ (although the latter case is unrealistic as this corresponds to purely radial orbits) for a representative CSiBORG simulation (9844) and rerun our analysis to obtain alternative constraints on a_1 .

For $\beta = 1$, one finds that the J factor diverges logarithmically towards the centre, and thus simply computing Eq. (4.41) would predict an infinite annihilation flux from each halo. In practice, if the annihilation rate becomes sufficiently high, then it will disrupt the density profile of the halo as the dark matter particles will be depleted towards the centre, presumably forming a cored profile. Thus, Eq. (4.41) is only valid up to some core radius, r_c . We estimate this to be the radius at which the dynamical time scale of the halo ($1/\sqrt{G\rho}$) equals the annihilation time scale ($\rho\sigma v/m_\chi$). Since we expect this to occur close to the centre of the halo, we use the leading order term for σ_r^2 ,

$$\frac{\sigma_r^2}{V_{\text{vir}}^2} \stackrel{\beta=1}{=} \frac{\pi^2 - 9}{6 \left(\log(1+c) - \frac{c}{1+c} \right) s} + \mathcal{O}(s^0), \quad (4.74)$$

and assume $\rho \sim \rho_0 r_s/r$ (see Eq. (4.44)). This leads to a core radius of

$$r_c \sim \left(\frac{\pi^2 - 9}{6} \frac{a_1}{m_\chi} \frac{V_{\text{vir}}^2}{\log(1+c) - \frac{c}{1+c}} \sqrt{\frac{\rho_0}{G}} \right)^{\frac{2}{3}} r_{\text{vir}}. \quad (4.75)$$

We have verified that, for the core radii that this implies, the $\mathcal{O}(s^0)$ term is negligible compared to the leading order term.

For simplicity, we compute Eq. (4.41) using r_c as the lower limit of the integral and add the contribution from the cored region separately. Furthermore, we note that Eq. (4.75) depends on a_1 and m_χ , and thus a fully-consistent analysis should recompute the J

factor at each step in the MCMC and for each channel. However, this is computationally infeasible. Instead, we set $a_1 = 2.4 \times 10^{-21} \text{ cm}^3 \text{ s}^{-1}$ and $m_\chi = 100 \text{ GeV}/c^2$ in this formula; this is a characteristic WIMP mass with a_1 at approximately the 20th percentile for the $b\bar{b}$ channel when $\beta = 0.5$. This typically leads to a core radius of $r_c \sim 0.1 - 1 \text{ pc}$. We find that the constraints on a_1 for $\beta = 1$ are much tighter than this, and thus our approximation yields a conservative constraint.

For both $\beta = 0.5$ and $\beta = 1$, we find that the anisotropic J factors lead to tighter constraints for all masses and channels, by up to a factor of 2.3 for $\beta = 0.5$ and up to two orders of magnitude for $\beta = 1$. This simply reflects the general behaviour of the J factor for these two different anisotropic profiles. As can be seen from Fig. 4.3, the $\beta = 0.5$ and $\beta = 1.0$ profiles are more strongly peaked towards the centre than the corresponding $\beta = 0$ case, with $\beta = 1.0$ diverging more quickly towards the centre and giving total J factors between 50 – 100 times larger than the corresponding $\beta = 0$ case. From this and Eq. (4.10), it is then easy to see that the derived constraints for these respective cases will show similar trends. Choosing $\beta = 0$ is therefore the most conservative of the cases considered, justifying our use of it in our fiducial analysis.

For all halos considered in this work, we have not attempted to add corrections to the halo density profiles due to baryonic effects, and have instead used profiles calibrated against dark-matter-only simulations. In the presence of baryons, adiabatic contraction during galaxy formation [238, 239] can steepen the central density profile, or it can be made shallower due to the subsequent stellar feedback [240, 241]. The different resulting density profiles would change our J factor calculations and hence our constraints on a_ℓ . Indeed, cosmological zoom hydrodynamical simulations have demonstrated that baryons do affect the J factors for velocity-dependent annihilation, typically increasing J from a factor of a few up to factors of several hundred [242, 243]. Assuming this increase in J can be applied to our halos, then our constraints on a_ℓ are conservative, however we leave it to future work to implement more precise corrections for baryonic physics.

4.5.2 Constrained simulation volume

The gamma rays emitted from either DM annihilation or decay would interact with the extragalactic background light (EBL) or CMB photons [244, 245]. This interaction manifests itself through pair production and therefore can cause signal attenuation, which can be modelled through an energy and redshift dependent optical depth coefficient $\tau(E, z)$. Since both the EBL and the CMB are approximately isotropic, the optical depth will not have a directional dependence. For the redshift range considered in this work ($z \lesssim 0.05$), the attenuation of the photon flux due to interaction with background photons will not be significant, except at very high energies ($\sim \text{TeV}$), which lie well above the maximum photon energies we consider here ($\sim 50 \text{ GeV}$), and thus we neglected this contribution.

Although this is the case for the very nearby Universe, there is also a contribution

to J and D from sources outside the CSiBORG volume. The expected contribution to the differential photon flux from s -wave annihilation is [246]

$$\begin{aligned} \left\langle \frac{d^2\Phi^{\text{ann}}}{dE_\gamma d\Omega} \right\rangle &= \frac{\langle\sigma v\rangle \bar{\rho}_{\text{DM},0}^2}{8\pi m_\chi^2} \int dz \left(\frac{dN_\gamma}{dE'_\gamma} \right) \Big|_{E'_\gamma=E_\gamma(1+z)} \\ &\times \frac{(1+z)^3}{H(z)} e^{-\tau(E'_\gamma,z)} \langle (1 + \delta(z, \Omega))^2 \rangle, \end{aligned} \quad (4.76)$$

where $\bar{\rho}_{\text{DM},0}$ represents present-day DM density and $\delta(z, \Omega)$ is the density fluctuation. This can be directly computed from the nonlinear matter power-spectrum (see, for example, [247]) or by using the halo model approach [177, 248]. Within the halo model, this factor comes directly from averaging the one-halo annihilation luminosity over the halo mass function. This is equivalent to marginalising over plausible realisations of the DM distribution in our Universe by utilising the Press-Schechter [249] formalism, or any other halo-formation model.

As in [248] (see Fig. 10 in their Appendix B), we estimate the integrand at a given E_γ and integrate between $z = 0.05$ and $z = 10$ to determine the ratio of this contribution to that explicitly modelled from CSiBORG. We find that this ratio is approximately unity at $E_\gamma = 5 \text{ GeV}$ for $m_\chi = 10 \text{ GeV}/c^2$ for the $b\bar{b}$ channel. One may be concerned that this is an important contribution; however, since our chosen HEALPIX resolution of `nside` = 256 corresponds to a physical scale of $\sim 0.6 h^{-1} \text{ Mpc}$ at the edge of the CSiBORG volume, one would expect that the extragalactic sources beyond $z \gtrsim 0.05$ are unresolved, and therefore this contribution will almost entirely be absorbed into the isotropic template. Of course, clustering of sources at redshifts beyond the CSiBORG volume would lead to an anisotropy in this unresolved emission. Our constraints are completely independent of how one models the isotropic part of the high-redshift component, and we leave it to further work to model the fluctuations about this, for example, by including constrained realisations of the density field for larger volumes. Since this contribution can only increase the J factor, we always underestimate the J factor in our templates, making our constraints conservative.

We note that in Eq. (4.76) one must correct for the redshift of emission; i.e., the spectrum should be evaluated at $E_\gamma(1+z)$ for a source at redshift z if we observe a photon at energy E_γ . Since we only considered sources at $z \lesssim 0.05$, we neglected this effect. If we consider the extreme case where all our sources were actually at $z = 0.05$, we would find that our estimate of the flux at a given $\langle\sigma v\rangle$ is correct to within 5% in the five lowest energy bins for the $b\bar{b}$ channel at $m_\chi = 100 \text{ GeV}/c^2$ (similar effects are seen for other channels and masses). Since this is comparable to the size of other reasonable variations to the model and this is an unrealistically extreme case, we are justified in making this assumption.

The question regarding the sufficiency of our CSiBORG volume also affects the constraints derived on p - and d -wave channels. Given that the most massive halos in the

local universe dominate the p - and d -wave constraints (running the inference using the ~ 20 largest halos gives constraints which differ by a factor only $\mathcal{O}(1)$ compared to the full result) raises the question of whether stronger constraints could be obtained using a dedicated cluster catalogue extending to higher redshift than CSiBORG. To address this we calculate the $J^{(\ell>0)}$ factor produced by the Master Catalogue of X-ray Clusters (MCXC; [250]), a compilation of clusters detected through their X-ray-emitting gas.

We find that the total $J^{(\ell>0)}$ factor of the CSiBORG catalogue is larger by factor of ~ 10 than that from MCXC. This is a result of the incompleteness of the MCXC catalogue as well as the various selection effects involved, which outweigh its larger redshift range and are absent from CSiBORG. The CSiBORG suite is therefore optimal in terms of its all-sky constraining power.

Nevertheless, since the p - and d -wave annihilation channels are strongly dependent on relative velocity of dark matter particles, and thus the total halo mass, it might still be worthwhile to explore constrained simulations with larger simulation volumes than CSiBORG to include the contributions from more distant massive halos. Further constraining power may be extracted from the angular dependence of the s -, p - and d -wave annihilation signals, which we do not fully exploit due to the relatively coarse resolution of our all-sky maps. Useful future work would therefore be to use higher-resolution simulations or observations of clusters to compare their predicted and observed flux profiles in greater detail. Note that the angular resolution of the *Fermi* data is approximately $n_{\text{side}} \approx 1024$, so significant improvements in resolution are possible. The positions and masses of local dark matter clusters may be found in the public CSiBORG halo catalogues [186].

4.5.3 Cluster masses

As demonstrated in Fig. 4.12 and indicated in [192], for p -wave and d -wave annihilation, the constraints on cross-section are dominated by the largest mass objects. It is therefore more important to correctly obtain cluster masses than it is when considering s -wave models (compare with Fig. 9 from [132]). In this work we have used the masses obtained from constrained N -body simulations, with initial conditions inferred using the BORG algorithm. This is an alternative to traditional estimates of cluster masses, which typically use one of the virial theorem [251], X-ray measurements [252], the Sunyaev Zel'dovich effect [253, 254] or weak lensing [255, 256]. These measurements tend to lead to significantly different measurements of masses for the same clusters, e.g. Fig. 3 [257], so it is not clear whether these traditional methods offer a more appropriate route to obtaining J factor maps as each method would produce very different constraints.

It has been found that the halo mass function in CSiBORG is higher than the average Λ CDM expectation at the massive end [185]. Although cosmic variance is high in this regime, this still represents a $\sim 2\sigma$ effect. To determine the effect of this potential mass discrepancy, we obtain a single new posterior sample of the initial conditions with a

different gravity model (using the COmoving Lagrangian Acceleration (COLA) method [258]), different time-stepping in the forward-model, a different method for generating power spectra (CLASS [259] instead of Eisenstein & Hu [260]) and a more robust likelihood [111], designed to remove the effects of unknown systematic uncertainties on scales larger than 1.4° . We then run a simulation with these initial conditions but otherwise identical to CSiBORG. This reconstruction produces 2- and 3-point statistics of the density field and a halo mass function which is closer to that expected from a random Λ CDM realisation. However, the re-simulated halo masses are systematically smaller than those observed and some objects (e.g. the Perseus cluster) are less prominent than in our fiducial reconstruction. These modifications to the older BORG algorithm are studied in more detail in [261].

Due to the variety of changes, this simulation is useful to test the robustness of our results to the reconstruction analysis. We produce J factor maps for this simulation and rerun the end-to-end inference. We find this new simulation produces weaker constraints than our fiducial analysis, with changes in the constraints of up to a factor of ~ 5 and ~ 12 for p - and d -wave, respectively. This is to be expected given the systematically smaller cluster masses obtained with the updated initial conditions and is much greater than the variation in constraints between CSiBORG realisations, which is typically $\sim 40 - 50\%$. We note that this variation is comparable to what one would obtain using measured clustered mass because different measurements give highly variable results [257]. It is the most significant uncertainty in our analysis.

4.5.4 Astrophysical templates

In our fiducial analysis, we use the latest galactic diffuse model provided by the Fermi Collaboration (g11_iem_v07). Although we find little degeneracy between the normalisation of the astrophysical templates and the amplitude of the J -factor maps (and hence a_ℓ constraints), we nonetheless re-run our inference using an older template (g11_iem_v02) to estimate the systematic effect of the choice of galactic template. We find that the constraints usually weaken slightly when using the older model, by up to 9% for s -wave, up to 24% for p -wave and up to 31% for d -wave, although at low m_χ the d -wave constraints can be up to 7% tighter with the older model. This variation is small compared to that caused by the cluster masses. We also find that the constraints on Γ can vary by $\sim 35\%$, while the inferred power-law index p changes by only ~ 0.01 in total, and hence is insensitive to the choice of the non-DM templates.

4.5.5 Thermal relics

In this work we have focused on dark matter particles of mass $2 - 500 \text{ GeV}/c^2$ since, if the thermally-average annihilation cross-section is $\langle\sigma v\rangle_{\text{th}} \approx 3 \times 10^{-26} \text{ cm}^3\text{s}^{-1}$ [60] and these particles are thermal relics, then the current abundance of dark matter is reproduced.

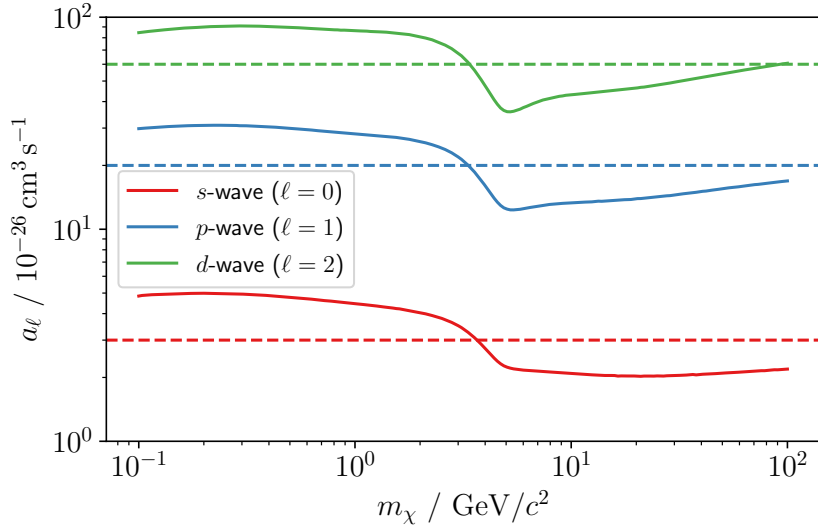


Figure 4.13: Annihilation cross-section coefficient ($\sigma v = a_\ell (v/c)^\ell$) as a function of dark matter particle mass, m_χ , required to give a present-day dark matter abundance of $\Omega_\chi h^2 = 0.11$ for a thermal relic. We plot the results for s -, p - and d -wave annihilation and see that the required a_ℓ increases by less than an order of magnitude for a unit change of ℓ . The dashed horizontal lines give the standard value of $a_0 = 3 \times 10^{-26} \text{ cm}^3 \text{ s}^{-1}$ for s -wave, and this value multiplied by $x_f/3$ and $x_f^2/20$ for p - and d -wave, assuming a freeze-out value of x of $x_f = 20$.

This cross-section is comparable to the electroweak coupling of the Standard Model, motivating the study of these dark matter candidates. In fact, the previously quoted $\langle \sigma v \rangle_{\text{th}}$ is appropriate for s -wave annihilation only. Hence, in this section we extend the analysis of [60] to obtain the equivalent values for p - and d -wave annihilation, and compare to the a_ℓ we obtain.

We assume that $\sigma v = a_\ell v^{2\ell}$ and that dark matter particles are non-relativistic and follow a Boltzmann distribution. The thermally-averaged cross-section at temperature T is then

$$\langle \sigma v \rangle = \frac{\int d^3 k_1 d^3 k_2 \exp\left(-\frac{E_1 + E_2}{T}\right) \sigma v}{\int d^3 k_1 d^3 k_2 \exp\left(-\frac{E_1 + E_2}{T}\right)} = \frac{4^\ell \Gamma\left(\ell + \frac{3}{2}\right) a_\ell}{\Gamma\left(\frac{3}{2}\right) x^\ell}, \quad (4.77)$$

where k_1 and k_2 are the phase-space momenta and $x = m_\chi/T$. One must then solve the equation governing the number density, n , of dark matter

$$\frac{dY}{dx} = \frac{s \langle \sigma v \rangle}{Hx} \left[1 + \frac{1}{3} \frac{d(\log g_s)}{d(\log T)} \right] (Y_{\text{eq}}^2 - Y^2), \quad (4.78)$$

where $Y = n/s$ for entropy density s , g_s is the number of relativistic degrees of freedom contributing to the entropy density, H is the Hubble parameter, and the equilibrium

value of Y is

$$Y_{\text{eq}} = \frac{45}{2\pi^4} \left(\frac{\pi}{8}\right)^{1/2} \frac{g_\chi}{g_s} x^{3/2} \exp(-x), \quad (4.79)$$

where the dark matter particle has g_χ degrees of freedom. Since this is a numerically stiff differential equation, we follow [60] and solve for $W = \log Y$ instead. We also take $g_s(T)$ from the calculations of [61]. Using an implicit Runge-Kutta method of the Radau IIA family of order 5 [262], we solve from $x = 1$ to $x = 10^3$, with an initial condition set using the approximate solution [263]

$$Y(x) \approx \frac{x^{\ell+2}}{2\lambda} + Y_{\text{eq}}(x), \quad (4.80)$$

which is valid long before freeze-out, where $\lambda \equiv 2.76 \times 10^{35} m_\chi \langle \sigma v \rangle_0$ for $\langle \sigma v \rangle = \langle \sigma v \rangle_0 x^{-\ell}$ and where m_χ is in GeV and $\langle \sigma v \rangle_0$ is in $\text{cm}^3 \text{s}^{-1}$. We then convert our solution to a density parameter

$$\Omega_\chi = \frac{8\pi G}{3H_0^2} m_\chi s_0 Y_0, \quad (4.81)$$

where H_0 , s_0 and Y_0 are the values of H , s and Y evaluated today. By comparing this value to $\Omega_\chi h^2 = 0.11$ [2, 3], for $H_0 = 100h \text{ km s}^{-1} \text{ Mpc}^{-1}$, we repeat the analysis for varying a_ℓ at fixed m_χ to obtain the thermal relic cross-section which reproduces the present-day dark matter abundance.

This is plotted in Fig. 4.13 as a function of m_χ for $\ell = 0, 1, 2$. We find that the required p -wave coefficient is $a_1 \sim (1 - 3) \times 10^{-25} \text{ cm}^3 \text{ s}^{-1}$, whereas for d -wave this is $a_2 \sim (4 - 9) \times 10^{-25} \text{ cm}^3 \text{ s}^{-1}$. In both cases, this is several orders of magnitude smaller than the constraints we find in Fig. 4.10, and thus we cannot rule out dark matter being a thermal relic with a velocity-dependent cross-section using large scale structure. These conclusions are consistent with those discussed in the context of the extragalactic gamma-ray background [264, 265]; a detection of a p -wave annihilation signal would suggest a non-thermal origin of the present-day dark matter abundance.

We note that approximate analytic solutions of Eq. (4.81) exist for p - and d -wave annihilation, e.g. [174, 266]. In these cases, if freeze-out occurs when $x = x_f$ and x_f is the same for all ℓ , then to obtain the p - and d -wave $\langle \sigma v \rangle_{\text{th}}$ from the s -wave analysis, one would multiply $a_{0,\text{th}}$ by $x_f/3 \approx 6.7$ and $x_f^2/20 \approx 20$, respectively, where we have used $x_f \approx 20$. These ratios give the dashed horizontal lines in Fig. 4.13, when multiplied by the fiducial value of $a_0 = 3 \times 10^{-26} \text{ cm}^3 \text{ s}^{-1}$.

4.5.6 Comparison to the literature

To enable a comparison between the constraints on $\langle \sigma v \rangle$ and Γ obtained in this work using large-scale structure and those from the literature, we now briefly summarise other methods for inferring these parameters and the results they produce.

Annihilation

The release of energy by annihilating DM has the potential to affect several observables over the Universe’s history. The earliest important observable is the element yield from Big Bang Nucleosynthesis (BBN): a $100 \text{ GeV}/c^2$ DM particle with the thermal relic cross section would release $\sim 1 \text{ MeV}$ of energy for every baryon in the Universe per Hubble time during the BBN era. This has the potential to alter subdominant nuclear reactions, although the effect is not strong enough to lead to stringent constraints [267, 268]. The next important epoch is recombination, where annihilating DM has the potential to ionise a non-negligible fraction of the hydrogen in the Universe. This would absorb CMB photons after recombination, to which the CMB angular power spectra are acutely sensitive [269]. This allows thermal relic DM with a velocity-independent cross section to be ruled out for masses below $\sim 10 - 30 \text{ GeV}/c^2$ depending on the annihilation channel [270, 271]. Of course, these bounds could be evaded by a large branching fraction into neutrinos or other particles with no electromagnetic interaction. The CMB constraints are particularly important for light DM ($m_\chi \lesssim 1 \text{ GeV}/c^2$) where the effective area and angular resolution of telescopes such as *Fermi*-LAT are poor. On the other hand, combining results from HST with BAO measurements, a joint analysis of data from Planck, WMAP9, ACT and SPT found no evidence of p -wave dark matter annihilation, ruling out cross-sections corresponding to $a_1 \sim 10^{-16} - 10^{-12} \text{ cm}^3 \text{ s}^{-1}$ for the production of leptons at masses $m_\chi \sim 1 - 10^3 \text{ GeV}/c^2$ at 95% confidence [272]. Our constraints are approximately four to five orders of magnitude more stringent than these results. One can similarly constrain dark matter annihilation via this process using the Lyman- α forest, but these constraints are also ~ 4 orders of magnitude weaker than those found for velocity-dependent annihilation in this work [273].

The above bounds were derived purely by considering the effects of energy injection into the Universe, but more information is available from observations of the potential annihilation products themselves. This is most often done by means of high-energy photons (a common final product regardless of annihilation channel), and forms the context for our own analysis. Of course, the spectra of the final-state photons depend crucially on the channel, as we have described previously. In the local Universe, the most promising targets are the galactic centre, nearby groups or clusters, and dwarf galaxies in the Local Group. The former is the greatest nearby concentration of DM, but also suffers from large astrophysical backgrounds, and the expected signal depends sensitively on the poorly known DM density profile of the Milky Way. Nevertheless, there are claims for a gamma-ray excess that could be due to annihilating DM [137]. In particular, authors of [137] claim the excess is well fit for $m_\chi \sim 7 - 10 \text{ GeV}/c^2$ with $\langle\sigma v\rangle \sim (0.5 - 5) \times 10^{-26} \text{ cm}^3 \text{ s}^{-1}$ annihilating via the $\tau^+\tau^-$ channel. Although we cannot rule out the lower values of $\langle\sigma v\rangle$ for this mass range, we do find that $\langle\sigma v\rangle < 4.5 \times 10^{-26} \text{ cm}^3 \text{ s}^{-1}$ at 95% confidence for these masses and this channel, which is incompatible with the larger values of $\langle\sigma v\rangle$ reported.

Furthermore, combining the relic abundance requirements with the GCE and lack of detection in dwarf galaxies can also be used to put DM annihilation constraints, as proposed by [274]. They suggested that a metastable particle undergoing s -wave annihilation can produce the thermal abundance of dark matter particles, and these daughter particles can then undergo p -wave annihilation in the present epoch. They find that a dark matter particle mass $m_\chi \sim 90 \text{ GeV}/c^2$ with an annihilation cross-section given by $a_1 \sim 10^{-20} \text{ cm}^3\text{s}^{-1}$ for the $b\bar{b}$ channel can reproduce the GCE and survive all other tests. For this mass and channel, the a_1 required to explain the GCE is ruled out at over 1σ confidence in our analysis, suggesting that large scale structure disfavours this interpretation. In particular, they convert constraints on $\langle\sigma v\rangle$ from clusters to constraints on a_1 by multiplying by multiples of the square of the velocity dispersion, and they find the Fornax cluster to be the most constraining object, with constraints which are ~ 2 orders of magnitude weaker than those given in Fig. 4.10. Since we perform a more careful analysis and obtain tight constraints on a_1 , future work should be dedicated to repeating the analyses cited by [274] but with the correct velocity-dependent J factors.

One can also use cosmic ray data in constraining DM annihilation. If dark matter annihilated via the e^+e^- channel, then the resulting cosmic rays from the Milky Way could be detectable in Voyager 1 and the AMS-02 data. [233] considered this effect, and found constraints on p -wave annihilation for this channel which are tighter than our constraints, by up to approximately one order of magnitude (see Fig. 4.10). Their results are found to be robust to choice of profile used for the galactic halo, although are only applicable to the e^+e^- channel. This is supported by [275], who obtain a null detection of p -wave annihilation as the cause of the GCE, since they do not find a spike in emission around Sgr A*. Moreover, since s - and p -wave annihilation produce different spatial distributions of emission [216] for the same density profile, it is argued that, to explain the GCE, p -wave annihilation would require a slope of the density profile of the Milky Way much steeper than normally considered ($\gamma \sim 1.5 - 1.7$ compared to $\gamma \sim 1.2 - 1.4$ for s -wave) [276], further questioning whether velocity-dependent annihilation is the cause of this detection.

Another promising laboratory for probing annihilation constraints are clusters of galaxies. They offer a massive accumulations of DM and permit a statistical analysis. However, they also suffer from potentially significant backgrounds. Dwarf galaxies, although smaller and less dense, have a much lower baryonic mass and hence the lowest contribution from degenerate astrophysical effects, affording a cleaner test. However, one is limited to a small sample size and thus one has to assume that the objects do not have peculiarities, e.g. unusual boost factors. By looking at a larger number of sources, as we do here, one can average over a more representative sample of substructure.

Fermi-LAT has been used to set limits on the annihilation cross section using dwarf galaxies, the Milky Way halo [182, 277, 278], and galaxy groups [279]. The strongest constraints come from the dwarfs, which, due to their lower distances, offer higher peak

signals than clusters [280]. These have been used to rule out the thermal relic cross section for masses below $\sim 100 \text{ GeV}/c^2$ assuming annihilation to b quarks [168, 169], as depicted in Fig. 4.7 (although see [281]). Even stronger constraints, ruling out the thermal relic scenario to $\mathcal{O}(\text{TeV})$ mass scales for annihilation to $b\bar{b}$, have been claimed from a radio search of the Large Magellanic Cloud [282]. Somewhat weaker constraints have also been obtained using dwarf irregular galaxies [170] and by cross-correlating *Fermi*-LAT data with the positions of nearby galaxies without knowledge of those galaxies' distances (dashed line in Fig. 4.7) [229, 283]. Further information can be gleaned by cross-correlating gamma-ray flux with a tracer of density such as gravitational lensing [284].

Similarly, in Fig. 4.10 we compare our p -wave constraints to those obtained by dwarf galaxies [231, 232, 285] for the $b\bar{b}$ channel, and find that our constraints are up to two orders of magnitude tighter. We also find that our d -wave constraints are approximately seven orders of magnitude tighter than those previously found from studying 25 dwarf spheroidal galaxies of the Milky Way [232], and hence lie outside the range of Fig. 4.10. This is in contrast to the results of s -wave annihilation, where dwarf galaxies give a constraint which is an order of magnitude tighter than that from large scale structure [132]. Assuming that clusters and dwarf galaxies have velocity dispersions of $\mathcal{O}(100) \text{ km s}^{-1}$ and $\mathcal{O}(1 - 10) \text{ km s}^{-1}$, respectively, then one could estimate the p - and d -wave constraints by multiplying the s -wave ones by $\sigma^{2\ell}$. Given this ratio of velocity dispersions, one would anticipate that clusters can produce p - and d -wave constraints which are factors of 10^{2-3} and 10^{5-6} tighter, approximately as we find. Despite being less constraining for velocity-dependent annihilation, if one did detect such a signal, then dwarf galaxies could still have an important role, since each object would have the same spectrum, but the different density and velocity profiles would result in different spatial distributions of the signal, which one could use to determine ℓ [286].

Decay

Similar considerations to those of the above subsection allow cosmological constraints to be placed on DM decay. These constraints are stronger at lower redshifts, where a greater fraction of DM decays per Hubble time for fixed decay rate. This allows BBN to test decay lifetimes around 10^{18}s , the CMB 10^{25}s , and the Lyman- α forest $10^{25} - 10^{26}\text{s}$ [246].

Constraints can also be derived from gamma-ray and neutrino telescopes. In particular, data from *Fermi*-LAT, AMS-02, PAO, KASCADE, and CASAMIA have been used to constrain the DM lifetime at the $10^{27} - 10^{28}\text{s}$ level for $10^2 \text{ GeV}/c^2 < m_\chi < 10^{17} \text{ GeV}/c^2$ [287, 288]. For lower-mass DM decaying primarily leptonically, bounds at the $10^{25} - 10^{26}\text{s}$ level can be set from X-ray and gamma-ray telescopes, the spectrometer on board the Voyager I spacecraft, and the heating of gas-rich dwarf galaxies, as well as the Lyman- α forest and CMB as described above [273, 289]. These constraints imply that over a very

large DM mass range, only a tiny fraction of the total DM can decay during the lifetime of the Universe. Decaying DM can also be constrained using the masses and abundances of Milky Way satellites in case the decay gives momentum to the DM particle, which provides a constraint of order the age of the Universe (e.g., [290]). The inferred values of Γ in this work are compatible with these constraints.

4.6 Conclusions

Annihilation of dark matter particles with velocity-dependent cross-sections can result in fluxes of gamma rays which are higher from extragalactic halos than from dwarf galaxies in the Local Group. In this work, we use the CSiBORG suite of constrained N -body simulations of the nearby Universe (within $155 h^{-1}\text{Mpc}$) to forward-model the expected emission from all halos down to $M_{\text{vir}} = 4.4 \times 10^{11} M_{\odot}$ and therefore produce full-sky, field-level, energy-dependent intensity predictions for s -, p -wave and d -wave dark matter annihilation as well as DM decay. We combine these with templates describing emission by baryonic processes in the Milky Way, an isotropic background and point sources, and compare to observations from the *Fermi* Large Area Telescope. Marginalising over the amplitudes of these templates and uncertainties in the reconstructed dark matter density field, we find no evidence for velocity-dependent dark matter annihilation to any standard model particle.

In case of s -wave annihilation, we rule out the thermal relic cross section at 95% confidence for DM particles of mass $m_{\chi} \lesssim 7 \text{ GeV}/c^2$ whose annihilation produces gluons or quarks less massive than the bottom quark. Our constraints for the production of charged leptons are approximately an order of magnitude less stringent, and we are unable to rule out the thermal relic cross section for the production of top or bottom quarks in our fiducial analysis. We infer at 3.3σ a nonzero contribution to the gamma-ray sky with the same spatial distribution as predicted by DM decay. For the decay to quarks, this corresponds to a decay rate of $\Gamma \approx 6 \times 10^{-28} \text{ s}^{-1}$. However, we find that a power-law spectrum is preferred by the data, and we infer that the power-law index is $p = -2.75_{-0.46}^{+0.71}$. If we marginalise over the contribution with the same spatial distribution as DM decay, we obtain constraints on $\langle\sigma v\rangle$ which are twice as tight as our fiducial analysis.

We also place upper limits for velocity-dependent annihilation channels. For DM mass in the range $m_{\chi} = 2 - 500 \text{ GeV}/c^2$, for the $b\bar{b}$ channel, defining $\sigma v = a_{\ell} (v/c)^{2\ell}$ where $\ell = 1, 2$ for p and d -wave, respectively, we constrain $a_1 < 2.4 \times 10^{-21} \text{ cm}^3 \text{ s}^{-1}$ and $a_2 < 3.0 \times 10^{-18} \text{ cm}^3 \text{ s}^{-1}$ at 95% confidence for $m_{\chi} = 10 \text{ GeV}/c^2$. Our constraints are weaker for the e^+e^- channel, with $a_1 < 1.5 \times 10^{-20} \text{ cm}^3 \text{ s}^{-1}$ and $a_2 < 1.9 \times 10^{-17} \text{ cm}^3 \text{ s}^{-1}$ at 95% confidence for this particle mass. We find that these constraints are dominated by the most massive halos ($M_{\text{h}} \sim 10^{14-16} M_{\odot}$). Our constraints are four to five orders of magnitude more stringent than those on p -wave annihilation from cosmic microwave

background analyses, and two (seven) orders of magnitude tighter than p -wave (d -wave) constraints from dwarf galaxies, as one would expect from the larger velocity dispersion in higher mass objects. Our bounds are however still several orders of magnitude above the value required to rule out dark matter being a thermal relic with a velocity-dependent annihilation cross-section.

In this work we have demonstrated the constraining power of the full halo field on velocity-dependent dark matter annihilation through a full-sky field-level inference, and obtained constraints which are several orders of magnitude tighter than alternative probes. Future work should be dedicated to higher-resolution inferences using the most constraining (highest mass) objects, incorporating other observations to constrain and marginalise over the density and velocity profiles, as well as investigating the case of Sommerfeld-enhanced annihilation channels as proposed in [190, 191].

The dominant systematic uncertainty in our analysis is the density field reconstruction (in particular its ability to reproduce high halo masses accurately), and hence future work should focus on understanding and improving this modelling (see [261]). Since anisotropic motions and baryonic processes can increase the J factor and hence the predicted signal, more precise modelling of these effects may achieve tighter constraints even with the same data.

4.7 Appendix A: Calculation of J factor for d -wave

To calculate the d -wave J factor at a given distance from the halo centre, one needs to evaluate the corresponding fourth velocity moment at that distance, $\langle \mathbf{v}^4 \rangle$. To do so, we assume ergodicity of the halo dark matter distribution function, which directly leads to the relation Eq. (4.49). However, this equation solves for the fourth moment of the radial velocity component, while we need the total $\langle \mathbf{v}^4 \rangle$. One can show that in the case of an ergodic distribution function the following holds (see Problem 4.29 in [221] and Section 1.3 from [291])

$$\langle v_\theta^{n-j} v_\phi^{j-k} v_r^k \rangle = \langle v_r^n \rangle \frac{\Gamma(\frac{k+1}{2}) \Gamma(\frac{n-j+1}{2}) \Gamma(\frac{j-k+1}{2})}{\pi \Gamma(\frac{n+1}{2})}. \quad (4.82)$$

From this, the following relations directly follow

$$\langle v_\theta^4 \rangle = \langle v_\phi^4 \rangle = \langle v_r^4 \rangle \quad (4.83)$$

$$\langle v_\theta^2 v_\phi^2 \rangle, = \langle v_\theta^2 v_r^2 \rangle = \langle v_\phi^2 v_r^2 \rangle = \frac{2}{3} \langle v_r^4 \rangle, \quad (4.84)$$

as does Eq. (4.50). Now, the only remaining issue is finding a solution to Eq. (4.49), since the second part of Eq. (4.48) can be directly obtained from Eq. (4.45). Note, that since we are assuming ergodicity, $\beta = 0$.

We assume NFW profiles (Eq. (4.44)) for our halos, which we rewrite as

$$\rho(s) = \frac{\rho_0}{cs(1+cs)^2}, \quad (4.85)$$

where c is the concentration, and $s = r/r_{\text{vir}}$. We can then solve Eq. (4.49) using Eq. (4.45). Noting that the gravitational potential in this notation is

$$\Phi(s) = -4\pi G\rho_0 r_s^2 \frac{\log(1+cs)}{cs}, \quad (4.86)$$

one obtains

$$\begin{aligned} \rho\langle v_r^4 \rangle &= \frac{4}{3}\pi^2 G^2 \rho_0^3 r_s^4 \times \\ &\times \left[-\frac{6(cs(cs(8cs+21)+6)-1)\log^2(cs+1)}{c^3 s^3} + \frac{6\text{Li}_2(c^2 s^2)(10cs+9\log(cs+1))}{cs} \right. \\ &\quad - \frac{6\log(cs+1)}{c^2 s^2 (cs+1)} \left(cs(-3\pi^2(cs+1)+cs(3cs+35)+25) + 3cs(cs+1)\log(cs)+2 \right) \\ &\quad + 288\text{Li}_3(cs+1) \\ &\quad + \text{Li}_2(cs) \left(-\frac{108\log(cs+1)}{cs} - 120 \right) + 72\text{Li}_2(cs+1)(3-4\log(cs+1)) \\ &\quad - \frac{52(\pi cs + \pi)^2 + 174cs + 177}{(cs+1)^2} + \frac{6}{cs(cs+1)^2} \\ &\quad \left. + \frac{6(8cs+9)\log^3(cs+1)}{cs} + 72\log(cs)(3-2\log(cs+1))\log(cs+1) + 18\log(cs) \right] \\ &+ \text{const}, \end{aligned} \quad (4.87)$$

where Li_2 and Li_3 represent the dilogarithm and trilogarithm, respectively. We use the implementations provided in the `polylogarithm` library⁷ to evaluate these. We impose $\lim_{s \rightarrow \infty} (\rho\langle v_r^4 \rangle)(s) = 0$, which fixes the integration constant to 0. This is a consequence of the fact that the velocity dispersion itself goes to zero far away from the halo, for the case of ergodic phase-space distribution function which we assumed throughout for our halos. This in turn implies that the velocity distribution has a vanishing second moment and hence, by definition, the fourth moment is zero too in this limit.

This analytical solution was implemented within the `CLUMPY` package and tested against a numerical solution. For all our results regarding d -wave, we always utilise the analytical solution from Eq. (4.87) and use the numerical solution only for cross-checking the results. We demonstrate the obtained J factor for d -wave in Fig. 4.3 obtained by

⁷<https://github.com/Expander/polylogarithm>

substituting the expression from Eq. (4.87) into Eq. (4.48) and normalising by the total integrated J factor value and dividing by the corresponding quantity for s -wave (labelled as $\tilde{J}^{\ell=2}/\tilde{J}^{\ell=0}$). One sees that the shape of the angular distribution of the normalised J factor is qualitatively very similar to the $\beta = 0$, p -wave case and shows the same behaviour towards the halo centre and the edge of the halo.

Note that similar results have been obtained in [216], who used Eddington inversion formula for obtaining the velocity moments. We chose to directly integrate the Jeans equations instead, since we were also interested in studying the anisotropic cases for the p -wave annihilation channel, and hence for consistency used the same approach for the d -wave J factor calculation.

CHAPTER 5

INFORMATION GAIN MAPS FROM THE LOCAL UNIVERSE

The following chapters first appeared in [292] with me as a first author and Jens Jasche, Doogesh Kodi Ramanah and Guilhem Lavaux as co-authors.

Abstract

We present maps classifying regions of the sky according to their information gain potential as quantified by the Fisher information. These maps can guide the optimal retrieval of relevant physical information with targeted cosmological searches. Specifically, we calculate the response of observed cosmic structures to perturbative changes in the cosmological model and chart their respective contributions to the Fisher information. Our physical forward modeling machinery transcends the limitations of contemporary analyses based on statistical summaries to yield detailed characterizations of individual 3D structures. We demonstrate this using galaxy counts data and showcase the potential of our approach by studying the information gain of the Coma cluster. We find that regions in the vicinity of the filaments and cluster core, where mass accretion ensues from gravitational infall, are the most informative about our physical model of structure formation in the Universe. Hence, collecting data in those regions would be most optimal for testing our model predictions. The results presented in this work are the first of their kind and elucidate the inhomogeneous distribution of cosmological information in the Universe. This study paves a new way forward to perform efficient targeted searches for the fundamental physics of the Universe, where search strategies are progressively refined with new cosmological data sets within an active learning framework.

5.1 Introduction

“*Where in the Universe should we look to learn something new?*” This is one of the most outstanding questions asked by researchers of all times. It reflects the wish to confirm our current physical understanding of the world and find new evidence to shift prevailing scientific paradigms. The latest advances in cosmological observations and data analysis now unlock the opportunity to perform machine-aided targeted searches for cosmological physics across the Universe. This has become feasible due to the availability of large-scale inferences that are informed by physics and causality of the cosmic large-scale structures and their evolution with time.

State-of-the-art cosmological surveys gather scientific information from tracers of the cosmic large-scale structures, such as galaxies, through large-area homogeneous scans of the sky. The historical origin of such search strategies is inherently related to the theoretical formulation of cosmology. Without any detailed knowledge about the actual spatial distribution of cosmic matter, theorists resorted to predicting the mean value of statistical summaries, constituting ensemble averages over different realizations of the unknown cosmic matter distribution. Until today, most cosmological analyses are performed on the basis of two-point statistics [293–295], although there are ongoing efforts to go beyond two- or three-point analyses. The initial need to use statistical summaries was twofold. First, the calculation of detailed matter field realizations by evaluating gravitational structure growth via simulations was costly and complex. Therefore, cosmological perturbation theory provided an efficient alternative to make predictions on the average statistical properties of the cosmic structures [296, 297]. Secondly, little was known about the particular realization of the cosmic matter distribution of our Universe. The best way forward was to test statements that would be true on average using the quantitative diagnostics provided by statistical summaries.

The above situation has witnessed a dramatic improvement in the era of modern cosmology. The abundance of cosmological observations and the availability of computational resources have now contributed significantly in constructing a much more refined picture of the actual spatial configuration of the cosmic structures beyond characterizations via statistical summaries. In particular, we have proposed the Bayesian physical forward modeling framework [19–23, 187] as a novel method enabling us to reconstruct the full 3D density field underlying observed galaxies in surveys with high fidelity. This machinery performs a causal inference that is informed by physics of the cosmic large-scale structures, their initial conditions and dynamical evolution. By characterizing the full 3D density field, our inference framework exploits the information on the phase distribution and higher-order statistics in data, which is inaccessible to traditional approaches. Although analyses relying on the use of the information content contained in the phase distribution of the large-scale structures have been proposed since a long time (see for e.g. [298, 299]), it was not propagated in the same manner as in the afore-

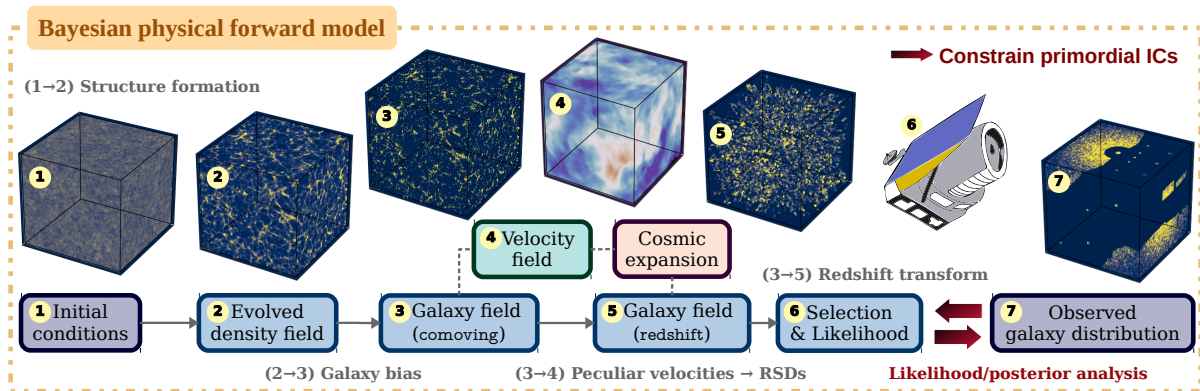


Figure 5.1: Schematic of the Bayesian physical forward modeling framework of BORG. BORG solves a large-scale Bayesian inverse problem by fitting a dynamical structure formation model to galaxy observations and subsequently inferring the primordial initial conditions (ICs) which lead to the formation of the presently observed cosmic structures via gravitational evolution (see also Section 2.6.1). The BORG forward modeling approach marginalizes naturally over unknown galaxy bias, and accounts for all relevant physical effects, such as the redshift space distortions (RSDs) resulting from the peculiar velocities of galaxies, and instrumental selection effects.

mentioned physical forward modeling machinery. Therefore, the inference carried out in previous studies by [20, 23, 187] transcends the limitations of contemporary analyses based on statistical summaries in order to yield detailed characterizations of individual 3D structures. Presently, apart from our framework, several research groups are developing the technology to perform full 3D characterizations of the cosmic structures probed by galaxy surveys [92, 94, 95, 300, 301].

So far, most of the work in the literature is focused on constraining cosmological physics from existing data sets, such as Cosmic Microwave Background (CMB) (see, for e.g., [5] and references therein) and large-scale structure, e.g. [295, 299, 302] observations. Here, we want to go beyond the standard task of constraining models to answering the question of how to optimally acquire future data that will be most informative to update our cosmological knowledge and to study physics. In doing so, we use existing information on cosmic structures to identify regions in the sky that promise the highest discovery potential as quantified by the Fisher information. Given the advent of the new data analysis technologies, we think that now it is especially timely to start a debate on whether the standard cosmological search strategy should be revised.

We note that some past studies aimed at understanding if certain regions of the sky are particularly informative about cosmology, e.g. [32], or if one could use such information to design better survey geometries in the future [33]. The method proposed

here goes beyond these studies by using the full forward model of the 3D matter field for quantifying the information content, weighted by the posterior of plausible realizations of the large-scale structures in the Universe.

In summary, we would like to address the following questions: Is it true that a homogeneous scan of the sky is the optimal search strategy to test fundamental physics with cosmological surveys? Are there some particular regions in the sky that are more informative than others with respect to specific research questions? Is it possible to answer some research questions faster with small, cheap and targeted searches? These questions are relevant, as they in turn raise questions on the optimal use of scientific resources. For example, two upcoming large-scale cosmological surveys, the Vera Rubin Observatory [303] and Euclid [304], constitute a total financial investment of over a billion dollars. As such, to optimize the scientific returns of these missions, we must ensure that the limited survey resources are adequately managed and optimally utilized.

With the ideas outlined in this work, we wish to trigger a discussion in the cosmological community on whether scientific progress in terms of information gain on fundamental physics can be sped up by using more refined cosmological searches, akin to active learning strategies. We also hope to initiate new technological developments for targeted searches of cosmological physics. To advance in this direction, here, we address the following question: “Which parts of the Universe should we observe to optimally retrieve information about a research question of interest?” Specifically, in this work, we are interested in identifying the regions in the Universe that are most relevant to obtain new information about the cosmological parameters underlying our standard model of cosmology. But our approach is equally applicable to other research questions of interest.

We present, for the first time, maps detailing the expected information gain provided by cosmic structures of the nearby Universe on cosmological parameters, based on reconstructions of the 2M++ galaxy catalog [305]. Our proposed methodology is, nevertheless, not limited to the nearby Universe and is applicable to any existing cosmological data sets.

5.2 Results

In our quest for mapping out the cosmologically sensitive large-scale structures in the sky, we adopt the Fisher information [306] methodology to provide a quantitative measure of the information encoded by the cosmological parameters on the observed large-scale structures. Although computing the Fisher information in this context is a highly non-trivial task, this ambitious undertaking is rendered feasible by employing a physical forward modeling machinery, such as our BORG (Bayesian Origin Reconstruction from Galaxies) algorithm [20, 23]. BORG constitutes a causal model of structure formation with a fully non-linear treatment of the dark matter clustering, allowing for the connection between the cosmic initial conditions and the observed galaxy distribution. A

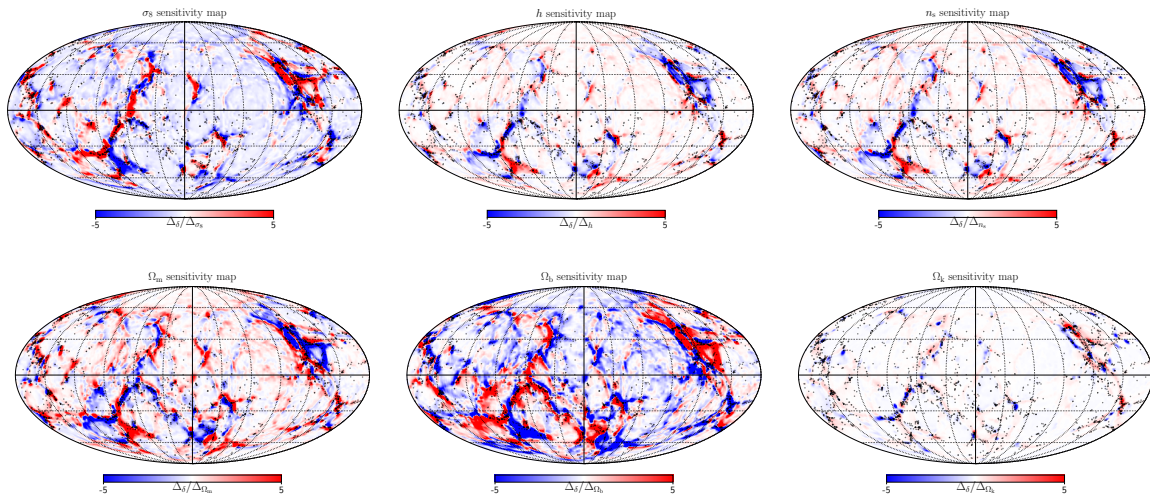


Figure 5.2: Components of the gradient of the matter density field with respect to the cosmological parameters for a spherical slice of thickness $\sim 2.65 h^{-1}\text{Mpc}$ at a comoving distance of $100 h^{-1}\text{Mpc}$ from the observer, with the corresponding galaxies from the 2M++ catalog denoted via black dots. Visually, we find that the baryon density Ω_b has the largest influence on the spatial distribution of the cosmic structures, with the regions surrounding the filamentary galaxy distribution being particularly sensitive to changes in the baryon density. Conversely, the least significant response emanates from the cosmic curvature Ω_k , with only the vicinity of the dense galaxy clusters reacting to changes in the geometry of the Universe.

schematic view of the distinct components of the BORG forward model is illustrated in Fig. 5.1. In our study, we marginalize over the ensemble of plausible realizations of the 3D primordial matter fluctuations of the very early Universe, as inferred by BORG via a hierarchical Bayesian statistical inference framework, from the 2M++ galaxy catalog [305] that traces the matter distribution of the nearby Universe. We refer the reader to the appendices for more detailed information pertaining to the BORG algorithm, 2M++ galaxy catalog, mathematical formalism underlying the Fisher information map and numerical implementation thereof. We stress that our approach for deriving the Fisher information map, based on the constrained realizations of cosmic structures conditional on galaxy observations, bears a stark contrast to the standard Fisher analyses to obtain forecasts on cosmological constraints from forthcoming galaxy surveys.

The availability of such inferred primordial matter fluctuations enables us to test the causal sensitivity of cosmic structures when forward modeled with perturbed cosmological parameters. To illustrate this possibility, in Fig. 5.2, we illustrate the response of cosmic structures in the Universe to perturbative changes in six parameters of the so-called concordance ΛCDM cosmological model, namely the matter density Ω_m , baryon

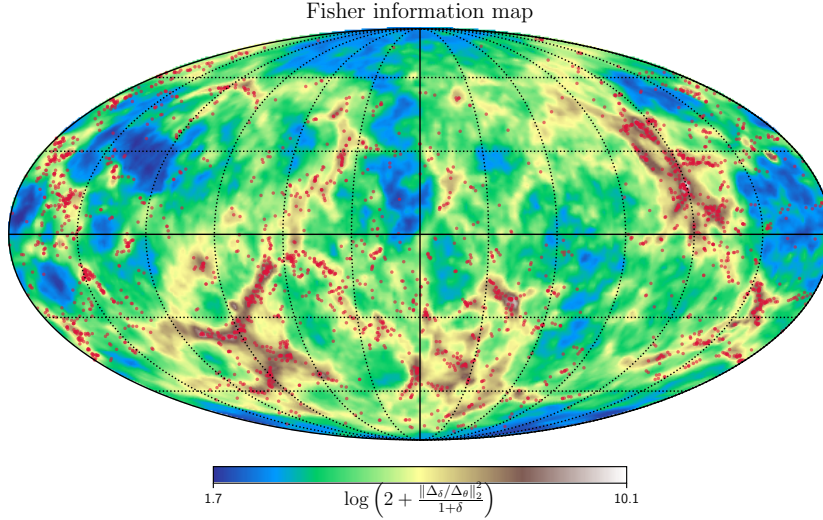


Figure 5.3: Fisher information map for the same spherical slice as in Fig. 5.2. The observed galaxy distribution from the 2M++ catalog, lying in the corresponding spherical shell centered around the observer, is represented by the red dots. We find that the regions in the vicinity of the massive cosmic structures, as traced by the galaxy distribution, are the most informative according to the Fisher information map. These regions correspond to the regime of gravitational infall of the galaxy clusters.

density Ω_b , cosmic curvature Ω_k , Hubble constant h , amplitude of matter fluctuations σ_8 and scalar spectral index n_s of the primordial power spectrum. This entails computing the gradient of the 3D galaxy field with respect to the cosmological parameters. This gradient quantifies the sensitivity of the galaxy distribution to changes in the cosmological parameters. As such, the gradient is described by a 6D vector for each volume element in the 3D grid. We compute the gradient using finite difference by executing BORG forward model evaluations on the ensemble of data constrained realizations of the cosmic initial conditions from the BORG 2M++ analysis [23], while varying the cosmological parameters about their corresponding fiducial values, as given by the latest best-fit values from the Planck Collaboration [5].

In Fig. 5.2, we visualize the individual gradient components of the six cosmological parameters by representing them as *cosmological sensitivity maps*. To this end, we consider a spherical slice of thickness $\sim 2.65 h^{-1}\text{Mpc}$ at a comoving distance of $100 h^{-1}\text{Mpc}$ from the observer that is projected onto a HEALPIX map [223]. The observed galaxies from the 2M++ catalog lying in the projected spherical slice are also indicated. From a visual comparison, we find that the density of baryonic matter, as characterized by the Ω_b parameter, induces the most significant response in the observed distribution of cosmic structures, with the Ω_m and σ_8 parameters also having a notable influence.

The astrophysical regions surrounding the filamentary patterns traced by the galaxies respond effectively to changes in the latter cosmological parameters. In contrast, the cosmic matter distribution displays a relatively minimal sensitivity to the Ω_k parameter, with the most noticeable effect of the cosmic curvature on the large-scale structures being on the densest regions, such as the galaxy clusters. The underdense regions of the matter distribution, the so-called cosmic voids, are most impacted by changes in the Ω_b and σ_8 parameters. We clarify that our structure formation model does not account for baryonic physics, so that Ω_b only modifies the amplitude of growth of matter perturbations and the shape of the linear power spectrum. Nevertheless, it is straightforward to repeat our analysis with hydrodynamical simulations to properly account for the presence of baryons. Given the significantly higher computational cost of such simulations, one possible option would be to use physics emulators, recently being developed on a large scale [307].

In addition to displaying the relative strength of the causal response of the cosmic structures to changes in our underlying cosmological model, the array of sensitivity maps in Fig. 5.2 reveals some interesting features. The σ_8 map, for example, illustrates how the clustering of matter occurs at the expense of the neighbouring regions, while the h map shows that a modified Hubble flow yields different clustering features. Moreover, the Ω_m and Ω_b maps display distinct anti correlated signatures. This may be attributed to the fact that the sound horizon distance scale due to baryon acoustic oscillations (BAOs) increases with both Ω_m and Ω_b , with the effect induced by Ω_b being much stronger, such that a fixed BAO scale encoded in the data yields this anti correlation between these two maps. Similarly, the Ω_m and σ_8 maps depict a striking anti correlation. This can be naturally explained on the basis that the primary constraining power emanates from the combination of the growth rate f of cosmological perturbations and σ_8 , where $f\sigma_8 \approx \sigma_8\Omega_m^{0.55}$ for the Λ CDM model.

A crucial ingredient in the Fisher information formalism derived in our study, as described by Eq. (5.12), lies in computing the above gradient. The desired 3D Fisher information field, obtained using Eq. (5.12), can be represented as a *Fisher information map* in the same way as the cosmological sensitivity maps. This Fisher information map represents the combined information gain on all six cosmological parameters considered in this study. The Fisher information map, for a spherical slice of thickness $\sim 2.65 h^{-1}\text{Mpc}$ at a distance of $100 h^{-1}\text{Mpc}$ from the observer, is displayed in Fig. 5.3, along with the observed galaxy distribution in this particular slice. The Fisher information map indicates that the regions with the highest information gain are those in the vicinity of the filamentary distribution traced by galaxies. One plausible interpretation is that the regions of gravitational infall, where there is the accumulation of matter, are the most informative about our cosmological model. This is particularly interesting as these infalling regions are, as yet, relatively obscure and must be properly understood.

The BORG analysis of the 2M++ galaxy catalog showcased the capacity of the physical

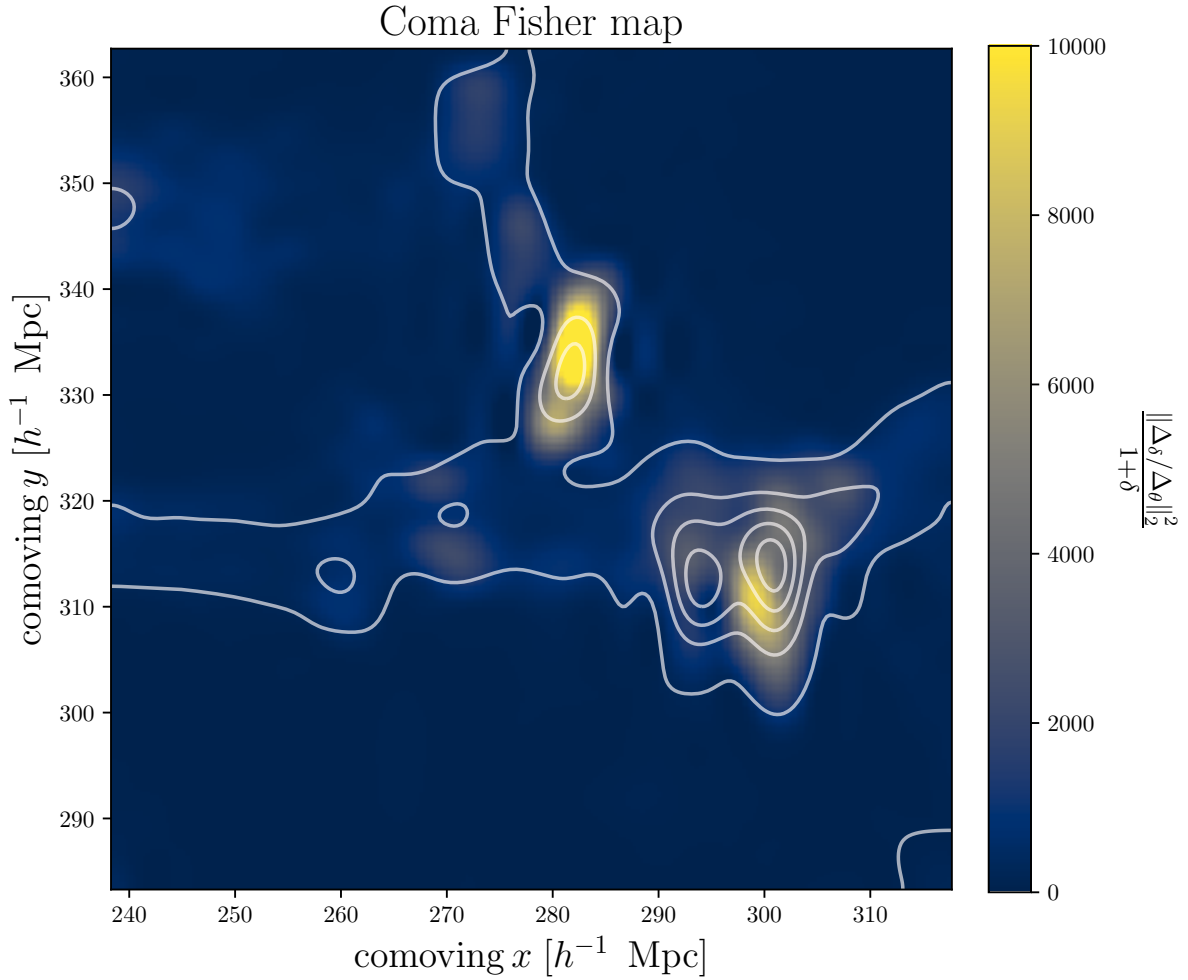


Figure 5.4: Fisher information map of the Coma cluster, with its corresponding mass density overlaid as a contour, for the central slice of thickness $\sim 5 h^{-1}$ Mpc through a 3D region centered on the cluster that extends over $40 h^{-1}$ Mpc. According to the Fisher information map, the central region containing the core of the cluster encodes fairly limited information gain, whilst the peripheral regions of the filaments and cluster core, where mass accretion occurs via gravitational infall, constitute the largest amount of information gain.

forward modeling machinery to resolve the key features of prominent cosmic structures in the present Universe [23]. The inferred mass profile of the Coma cluster, in particular, was found to be in remarkable agreement with state-of-the-art weak lensing measurements. Therefore, it is also interesting to study the characteristic features underlying the source of cosmological information for this well-known cluster. Fig. 5.4 displays the Fisher information map of the Coma cluster and its corresponding mass density. We note that the figure depicts the central slice of thickness $\sim 5 h^{-1}\text{Mpc}$ through a 3D patch extending over $40 h^{-1}\text{Mpc}$. The features present in the Fisher map of the Coma cluster are in accordance with those from the sky projected Fisher map from Fig. 5.3, thereby supporting the interpretation that the information that can be gleaned from the regions surrounding the filaments, i.e. the gravitationally infalling regions of the cluster, are the most significant. Recent studies, relying on conceptually distinct methodologies, also indicate that accretion filaments and the surroundings of voids are highly sensitive to predictions of dark energy [308] and gravity [309] models. In contrast, the central core of the cluster provides a relatively lower information gain, with the regions devoid of matter displaying unsubstantial information. A key point worth stressing here is that our relatively simple physics informed algorithm was capable of pinning down potential regions of cosmological interest in the sky.

Now, one can use the maps displayed in Figs. 5.3 and 5.4 as a guide towards where to look in order to optimally collect data for testing our physical model of structure formation. For the particular case demonstrated here, the data consists of galaxy counts and the corresponding information gain predictions are for the galaxy clustering data. The idea behind the targeted search approach is to recursively search for galaxies in the high information gain regions, as quantified by the Fisher map, then use this newly acquired data to constrain the model parameters and repeat the procedure. This is schematically depicted in Fig. 5.5. The justification pertaining to why this search strategy is optimal is further elaborated in Appendix 5.6. It should be understood that this approach can be similarly extended to other observables for constructing the corresponding information gain maps.

5.3 Discussion

The results presented in this work demonstrate the feasibility of machine aided targeted searches for cosmological physics signals. These have become feasible through recent developments of physics informed causal inference frameworks to study the 3D cosmic large-scale structures, their origin and evolution with time. We have used 3D initial conditions, inferred with the **BORG** algorithm, and a physics simulator of cosmological structure formation to chart the response of observed structures in the Universe and their corresponding information gain with respect to cosmological parameters. Our results are the first of their kind and elucidate the inhomogeneous distribution of cosmological infor-

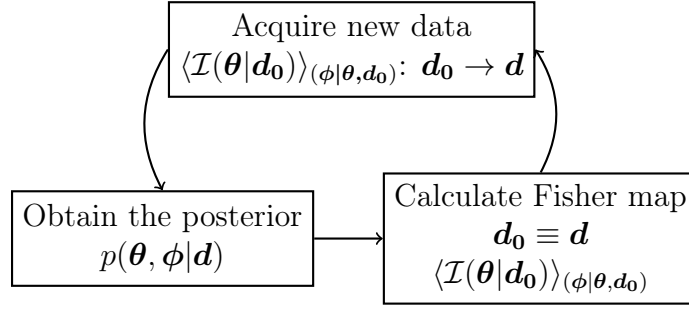


Figure 5.5: Flowchart of the targeted search approach. The core idea is to use the existing knowledge as quantified by the posterior $p(\boldsymbol{\theta}, \boldsymbol{\phi}|\mathbf{d})$ in order to calculate the Fisher map marginalized over this posterior $\langle \mathcal{I}(\boldsymbol{\theta}|\mathbf{d}_0) \rangle_{\langle \phi|\boldsymbol{\theta}, \mathbf{d}_0 \rangle}$, keeping the observables of interest. Then, this Fisher map will provide us with regions of the sky with highest information gain potential for acquiring new data optimal for testing our model predictions, $\langle \mathcal{I}(\boldsymbol{\theta}|\mathbf{d}_0) \rangle_{\langle \phi|\boldsymbol{\theta}, \mathbf{d}_0 \rangle}: \mathbf{d}_0 \rightarrow \mathbf{d}$. This data will in turn be used to update our knowledge about the model through updating the posterior and the procedure would repeat until the information content is fully depleted.

mation in the Universe. This study paves a new way forward to perform efficient targeted searches for the fundamental physics of the Universe, where search strategies are progressively refined with new cosmological data sets within an active learning framework.

We have further illuminated the response of the cosmic large-scale structures with respect to individual cosmological parameters. These results suggest that different features of the cosmic structures respond differently to perturbations in the physics determined by the cosmological parameters. For instance, we find that the vicinity of the filamentary cosmic structures, corresponding to gravitationally infalling regions, are highly sensitive to changes in the baryon density, with the cosmic curvature impacting only the surroundings of the dense galaxy clusters. Cosmic voids are primarily affected by the amplitude of matter fluctuations and the baryon density. Our results demonstrate the value of going beyond state-of-the-art analyses of the cosmic large-scale structures that are limited to summary statistics and ignore this richness of the 3D cosmic structures. Even though we consider one particular observable, namely galaxy counts, in this study, our proposed framework can be seamlessly applied to other tracers, such as Lyman- α forest [300, 310], which would bear complementary information.

Our findings further suggest that optimal targeted searches for research questions have become feasible. Given a specific research question, our approach enables us to propose targets for optimal information retrieval. This raises the question if traditional survey strategies should be revised and if significant scientific progress could also be driven by targeted searches with smaller, cheaper and faster instrumentation. We hope that our contribution will trigger a discussion and new technological advances in the

field.

While we studied the influence of cosmological parameters on the Universe in this work, our proof of concept reveals the potentially far-reaching and groundbreaking implications when considering the physical effects induced by modified gravity [311], dynamical dark energy [312] and exotic dark matter models, such as self-interacting [313] and fuzzy dark matter [314], or massive neutrinos [315], on the cosmic large-scale structures. Once we identify the astrophysical region(s) of interest, based on the Fisher information map, for a particular model, we may subsequently proceed by computing accurate theoretical predictions for the spatial distribution of matter and luminous tracers for the given model via cosmological N -body or hydrodynamical simulations with extremely high resolution, thereby resulting in highly detailed physical features. The final step would then entail observing the relevant region(s) and comparing the theoretical predictions with the galaxy observations in a likelihood or posterior analysis. In essence, we are proposing a novel way of doing science that optimizes scarce resources for maximal scientific returns via an efficient observational strategy.

5.4 Appendix A: 2M++ galaxy catalog

The 2M++ catalog [305] is a compilation galaxy redshifts derived from the Two-Micron-All-Sky-Survey (2MASS) Redshift Survey (2MRS) [316], the Six-Degree-Field Galaxy Redshift Survey Data Release 3 (6dFGRS-DR3) [317], and the Sloan Digital Sky Survey Data Release 7 (SDSS-DR7) [318]. The resulting catalog has a greater depth and a higher sampling rate relative to the Infrared Astronomical Satellite (IRAS) Point Source Catalog Redshift Survey (PSCZ) [319]. The photometry is based on the 2MASS Extended Source Catalog (2MASS-XSC) [320], an all-sky survey in the J , H and K_S bands, with redshifts in the K_S band of the 2MRS complemented by those from the SDSS-DR7 and 6dFGRS-DR3.

Since the 2M++ catalog is a combination of several surveys, the galaxy magnitudes from all sources were first recomputed by measuring the apparent magnitude in the K_S band within a circular isophote at 20 mags arcsec⁻². The apparent K_S band magnitudes were subsequently corrected by taking into account Galactic extinction, cosmological surface brightness dimming, stellar evolution and k -corrections, while masking the Galactic Plane. To account for the incompleteness due to fibre collisions in 6dFGRS and SDSS, the redshifts of nearby galaxies within each survey region are cloned. The final 2M++ catalog contains 69190 galaxies in total, and is fairly sampling the galaxies of the cosmic volume up to a distance of 200 h^{-1} Mpc for the area covered by the 6dFGRS and SDSS, and up to 125 h^{-1} Mpc for the region mapped by 2MRS. For a more in-depth description of the construction of the 2M++ catalog, we refer the interested reader to the original compilation [305], with the computation of radial selection functions and target selection completeness as required for the BORG 2M++ analysis [23] detailed in Section 4.1 thereof.

5.5 Appendix B: Fisher information maps

The proposed targeted search approach requires identifying regions of the observational domain, which, when observed, can provide optimal information gain to update existing knowledge. To quantify this information gain, we choose to employ Fisher information [306]. The latter measures the amount of information that unseen observable data can carry about uncertain model parameters. More explicitly, Fisher information estimates the squared norm of the likelihood score for given model parameters averaged over all possible future data realizations permitted by the likelihood. It, therefore, measures the expected strength with which the likelihood will respond to changes in the model parameters once new data becomes available. For the specific case considered in this work, we assume that prior knowledge on the spatial cosmic matter configuration is available from previous observations. More specifically, we assume that the white noise realizations (the phases) of the initial conditions ϕ are provided by the BORG reconstruction of the 2M++ survey [23]. Assuming this realization of initial conditions we can express the

conditional Fisher information as:

$$\begin{aligned} \mathcal{I}(\boldsymbol{\theta}|\boldsymbol{\phi}) &= \mathbb{E}_{(\mathbf{d}|\boldsymbol{\theta},\boldsymbol{\phi})} \left[\left(\frac{\partial \ln(\mathcal{L}(\mathbf{d}|\boldsymbol{\theta},\boldsymbol{\phi}))}{\partial \boldsymbol{\theta}} \right)^2 \right] \\ &= \int \mathcal{D}\mathbf{d} \left(\frac{\partial \ln(\mathcal{L}(\mathbf{d}|\boldsymbol{\theta},\boldsymbol{\phi}))}{\partial \boldsymbol{\theta}} \right)^2 \mathcal{L}(\mathbf{d}|\boldsymbol{\theta},\boldsymbol{\phi}), \end{aligned} \quad (5.1)$$

where $\boldsymbol{\theta}$ corresponds to the set of cosmological parameters parametrizing our forward model, with the galaxy observations within the voxels denoted by $\mathbf{d} = [N_i]_{i=1,\dots,N_{\text{box}}}$ and N_{box} corresponds to the number of the voxels inside the 3D volume considered in the 2M++ BORG reconstruction (see [23] for further details). Note that we express the vector nature of these random variables using boldface symbols. For the sake of this work, we assume a Poisson likelihood to describe the galaxy clustering data:

$$\mathcal{L}(\mathbf{d}|\boldsymbol{\theta},\boldsymbol{\phi}) = \prod_i e^{-\lambda_i(\boldsymbol{\theta};\boldsymbol{\phi})} \frac{\lambda_i(\boldsymbol{\theta};\boldsymbol{\phi})^{N_i}}{N_i!}, \quad (5.2)$$

where $\lambda_i(\boldsymbol{\theta};\boldsymbol{\phi})$ denotes the rate of the specific realization of the Poisson process as dictated by the initial phases $\boldsymbol{\phi}$ and cosmological parameters $\boldsymbol{\theta}$, while N_i represents the galaxy counts within the i -th cell of the 3D volume. Now, taking the derivative with respect to the cosmological parameters $\boldsymbol{\theta}$ gives:

$$\begin{aligned} \frac{\partial \ln(\mathcal{L}(\mathbf{d}|\boldsymbol{\theta},\boldsymbol{\phi}))}{\partial \boldsymbol{\theta}} &= \sum_i \left(-\frac{\partial \lambda_i(\boldsymbol{\theta};\boldsymbol{\phi})}{\partial \boldsymbol{\theta}} + \frac{N_i}{\lambda_i(\boldsymbol{\theta};\boldsymbol{\phi})} \frac{\partial \lambda_i(\boldsymbol{\theta};\boldsymbol{\phi})}{\partial \boldsymbol{\theta}} \right) \\ &= \sum_i \left(\frac{N_i}{\lambda_i(\boldsymbol{\theta};\boldsymbol{\phi})} - 1 \right) \frac{\partial \lambda_i(\boldsymbol{\theta};\boldsymbol{\phi})}{\partial \boldsymbol{\theta}}. \end{aligned} \quad (5.3)$$

The square of the above derivative is given by:

$$\begin{aligned} \left(\frac{\partial \ln(\mathcal{L}(\mathbf{d}|\boldsymbol{\theta},\boldsymbol{\phi}))}{\partial \boldsymbol{\theta}} \right)^2 &= \sum_{ij} \frac{\partial \lambda_i(\boldsymbol{\theta};\boldsymbol{\phi})}{\partial \boldsymbol{\theta}} \frac{\partial \lambda_j(\boldsymbol{\theta};\boldsymbol{\phi})}{\partial \boldsymbol{\theta}} \\ &\times \left(\frac{N_i}{\lambda_i(\boldsymbol{\theta};\boldsymbol{\phi})} \frac{N_j}{\lambda_j(\boldsymbol{\theta};\boldsymbol{\phi})} - \frac{N_i}{\lambda_i(\boldsymbol{\theta};\boldsymbol{\phi})} - \frac{N_j}{\lambda_j(\boldsymbol{\theta};\boldsymbol{\phi})} + 1 \right) \\ &= \sum_i \left(\frac{\partial \lambda_i(\boldsymbol{\theta};\boldsymbol{\phi})}{\partial \boldsymbol{\theta}} \right)^2 \left(\frac{N_i^2}{\lambda_i^2(\boldsymbol{\theta};\boldsymbol{\phi})} - 2 \frac{N_i}{\lambda_i(\boldsymbol{\theta};\boldsymbol{\phi})} + 1 \right) \\ &+ \sum_{ij} (1 - \delta_{ij}^K) \frac{\partial \lambda_i(\boldsymbol{\theta};\boldsymbol{\phi})}{\partial \boldsymbol{\theta}} \frac{\partial \lambda_j(\boldsymbol{\theta};\boldsymbol{\phi})}{\partial \boldsymbol{\theta}} \\ &\times \left(\frac{N_i}{\lambda_i(\boldsymbol{\theta};\boldsymbol{\phi})} \frac{N_j}{\lambda_j(\boldsymbol{\theta};\boldsymbol{\phi})} - \frac{N_i}{\lambda_i(\boldsymbol{\theta};\boldsymbol{\phi})} - \frac{N_j}{\lambda_j(\boldsymbol{\theta};\boldsymbol{\phi})} + 1 \right), \end{aligned} \quad (5.4)$$

where we split the sum in diagonal and off-diagonal terms. Again, the indices i and j denote the pixels in the HEALPIX projection. Since given a density field, the individual Poisson realizations are independent, the off-diagonal terms will vanish when computing the Fisher information as given by Eq. (5.1). In order to see this, consider the following data average of Eq. (5.4):

$$\begin{aligned}
\left\langle \left(\frac{\partial \ln(\mathcal{L}(\mathbf{d}|\boldsymbol{\theta}, \boldsymbol{\phi}))}{\partial \boldsymbol{\theta}} \right)^2 \right\rangle_{(\mathbf{d}|\boldsymbol{\theta}, \boldsymbol{\phi})} &= \sum_{ij} \frac{\partial \lambda_i(\boldsymbol{\theta}; \boldsymbol{\phi})}{\partial \boldsymbol{\theta}} \frac{\partial \lambda_j(\boldsymbol{\theta}; \boldsymbol{\phi})}{\partial \boldsymbol{\theta}} \\
&\times \left\langle \left(\frac{N_i}{\lambda_i(\boldsymbol{\theta}; \boldsymbol{\phi})} \frac{N_j}{\lambda_j(\boldsymbol{\theta}; \boldsymbol{\phi})} - \frac{N_i}{\lambda_i(\boldsymbol{\theta}; \boldsymbol{\phi})} - \frac{N_j}{\lambda_j(\boldsymbol{\theta}; \boldsymbol{\phi})} + 1 \right) \right\rangle_{(\mathbf{d}|\boldsymbol{\theta}, \boldsymbol{\phi})} \\
&= \sum_i \left(\frac{\partial \lambda_i(\boldsymbol{\theta}; \boldsymbol{\phi})}{\partial \boldsymbol{\theta}} \right)^2 \left(\frac{\lambda_i(\boldsymbol{\theta}; \boldsymbol{\phi})^2 + \lambda_i(\boldsymbol{\theta}; \boldsymbol{\phi})}{\lambda_i^2(\boldsymbol{\theta}; \boldsymbol{\phi})} - 2 \frac{\lambda_i(\boldsymbol{\theta}; \boldsymbol{\phi})}{\lambda_i(\boldsymbol{\theta}; \boldsymbol{\phi})} + 1 \right) \\
&\quad + \sum_{ij} (1 - \delta_{ij}^K) \frac{\partial \lambda_i(\boldsymbol{\theta}; \boldsymbol{\phi})}{\partial \boldsymbol{\theta}} \frac{\partial \lambda_j(\boldsymbol{\theta}; \boldsymbol{\phi})}{\partial \boldsymbol{\theta}} \\
&\quad \underbrace{\left(\frac{\lambda_i(\boldsymbol{\theta}; \boldsymbol{\phi}) \lambda_j(\boldsymbol{\theta}; \boldsymbol{\phi})}{\lambda_i(\boldsymbol{\theta}; \boldsymbol{\phi}) \lambda_j(\boldsymbol{\theta}; \boldsymbol{\phi})} - \frac{\lambda_i(\boldsymbol{\theta}; \boldsymbol{\phi})}{\lambda_i(\boldsymbol{\theta}; \boldsymbol{\phi})} - \frac{\lambda_j(\boldsymbol{\theta}; \boldsymbol{\phi})}{\lambda_j(\boldsymbol{\theta}; \boldsymbol{\phi})} + 1 \right)}_{=0}, \quad (5.5)
\end{aligned}$$

where we used the following identities:

$$\begin{aligned}
\langle N_i \rangle_{(\mathbf{d}|\boldsymbol{\theta}, \boldsymbol{\phi})} &= \lambda_i(\boldsymbol{\theta}; \boldsymbol{\phi}), \\
\langle N_i^2 \rangle_{(\mathbf{d}|\boldsymbol{\theta}, \boldsymbol{\phi})} &= \lambda_i(\boldsymbol{\theta}; \boldsymbol{\phi}) + \lambda_i^2(\boldsymbol{\theta}; \boldsymbol{\phi}) \quad \text{for } \forall i \in [1, N]. \quad (5.6)
\end{aligned}$$

The first equality follows from the conditional independence of the Poisson realizations within each grid cell, which always holds once the underlying density field $\boldsymbol{\lambda}(\boldsymbol{\theta}, \boldsymbol{\phi})$ is given, while the second results from the expression of variance of the Poisson distribution. Therefore, the diagonal terms are:

$$\mathcal{I}(\boldsymbol{\theta}|\boldsymbol{\phi}) = \sum_i \left(\frac{\partial \lambda_i(\boldsymbol{\theta}; \boldsymbol{\phi})}{\partial \boldsymbol{\theta}} \right)^2 \frac{1}{\lambda_i(\boldsymbol{\theta}; \boldsymbol{\phi})}. \quad (5.7)$$

From the above equation, we see that the contribution from a particular volume element to the Fisher information is given by:

$$h'_i(\boldsymbol{\theta}|\boldsymbol{\phi}) = \left(\frac{\partial \lambda_i(\boldsymbol{\theta}; \boldsymbol{\phi})}{\partial \boldsymbol{\theta}} \right)^2 \frac{1}{\lambda_i(\boldsymbol{\theta}; \boldsymbol{\phi})} \equiv \lambda_i(\boldsymbol{\theta}, \boldsymbol{\phi}) \left(\frac{\partial \ln \lambda_i(\boldsymbol{\theta}; \boldsymbol{\phi})}{\partial \boldsymbol{\theta}} \right)^2. \quad (5.8)$$

This allows us to obtain a 3D Fisher map. Now, we also have to specify the Poisson intensity in terms of the output of our physics simulator. This relation can be written as:

$$\lambda_i(\boldsymbol{\theta}; \boldsymbol{\phi}) = B(\delta \equiv G(\boldsymbol{\theta}; \boldsymbol{\phi})), \quad (5.9)$$

where $B(x)$ is an arbitrary non-negative bias function and $G(\boldsymbol{\theta}, \boldsymbol{\phi})$ is a physics simulator of the cosmic large-scale structures with output δ corresponding to the 3D matter density contrast amplitudes. For illustrative purposes, we assume $B(x) = 1 + x$. Note any bias model monotonic in the density will change the quantitative results, but not the qualitative results, as can also be seen from Eq. (5.8). We then obtain:

$$\lambda_i(\boldsymbol{\theta}; \boldsymbol{\phi}) = 1 + G_i(\boldsymbol{\theta}; \boldsymbol{\phi}), \quad (5.10)$$

which leads to the following expression for the derivative with respect to cosmological parameters $\boldsymbol{\theta}$:

$$\frac{\partial \lambda_i(\boldsymbol{\theta}; \boldsymbol{\phi})}{\partial \boldsymbol{\theta}} = \frac{\partial G_i(\boldsymbol{\theta}; \boldsymbol{\phi})}{\partial \boldsymbol{\theta}}. \quad (5.11)$$

We can therefore approximate the Fisher information elements through finite differencing as:

$$\begin{aligned} h'_i(\boldsymbol{\theta}; \boldsymbol{\phi}) &= \left(\frac{\partial G_i(\boldsymbol{\theta}; \boldsymbol{\phi})}{\partial \boldsymbol{\theta}} \right)^2 \frac{1}{1 + G_i(\boldsymbol{\theta}; \boldsymbol{\phi})} \\ &\approx \left(\frac{G_i(\boldsymbol{\theta}_0 + \Delta \boldsymbol{\theta}; \boldsymbol{\phi}) - G_i(\boldsymbol{\theta}_0; \boldsymbol{\phi})}{\Delta \boldsymbol{\theta}} \right)^2 \frac{1}{1 + G_i(\boldsymbol{\theta}_0; \boldsymbol{\phi})}. \end{aligned} \quad (5.12)$$

Note, however, that the above calculation is done for a fixed realization of the phases $\boldsymbol{\phi}$. Since we do not know what particular realization is compatible with our Universe, we need to marginalize over them:

$$\begin{aligned} h_i(\boldsymbol{\theta}|d) &= \int \mathcal{D}\boldsymbol{\phi} p(\boldsymbol{\phi}|d) h'_i(\boldsymbol{\theta}|\boldsymbol{\phi}) \\ &\equiv \int \mathcal{D}\boldsymbol{\phi} p(\boldsymbol{\phi}|d) \left(\frac{\partial G_i(\boldsymbol{\theta}; \boldsymbol{\phi})}{\partial \boldsymbol{\theta}} \right)^2 \frac{1}{1 + G_i(\boldsymbol{\theta}; \boldsymbol{\phi})}. \end{aligned} \quad (5.13)$$

This task is made possible by the BORG algorithm, as described in Section 2.6.1. In this way, we are able to use the information content of the previously obtained constraints and update the Fisher information. Projecting this 3D Fisher map onto a HEALPIX grid and visualizing a particular spherical slice through it, we obtain an all-sky map which we refer to in our study as the *Fisher information map*. A particular example is provided in Fig. 5.3.

Now, in order to calculate the $G_i(\boldsymbol{\theta}_0 + \Delta \boldsymbol{\theta}; \boldsymbol{\phi})$ term above, we first perform a forward model evaluation by setting the cosmological parameters to their corresponding fiducial values, $\boldsymbol{\theta}_0 \equiv \{\Omega_m = 0.3111, \Omega_b = 0.049, \Omega_k = 7 \times 10^{-4}, h = 0.6766, \sigma_8 = 0.8102, n_s = 0.9665\}$, according to the latest Planck best-fit Λ CDM cosmology [5], for all the MCMC realizations of the initial conditions from the BORG 2M++ analysis. We then compute the mean galaxy field marginalized over the forward model output realizations. We

subsequently repeat this procedure for a new set of perturbed cosmological parameters, $\boldsymbol{\theta}' \equiv \boldsymbol{\theta}_0 + \Delta\boldsymbol{\theta}$, ensuring $\Delta\boldsymbol{\theta}$ does not exceed values outside 1-sigma width of a Gaussian centered at the fiducial cosmology $\boldsymbol{\theta}_0$ as characterized by Planck best-fit values. Using the respective means of the fiducial and perturbed galaxy fields, we compute the gradient of the forward model output with respect to the cosmological parameters, $\Delta_G/\Delta\boldsymbol{\theta}$, using a finite differencing scheme, with:

$$\begin{aligned}\Delta_G &= \int \mathcal{D}\phi p(\phi|\mathbf{d}) [G(\boldsymbol{\theta}_0 + \Delta\boldsymbol{\theta}; \phi) - G(\boldsymbol{\theta}_0; \phi)] \\ &\approx \frac{1}{N_\phi} \sum_i [G(\boldsymbol{\theta}_0 + \Delta\boldsymbol{\theta}; \phi_i) - G(\boldsymbol{\theta}_0; \phi_i)] \quad \text{with } \phi_i \leftarrow p(\phi|\mathbf{d})\end{aligned}$$

and $\Delta\boldsymbol{\theta} = \boldsymbol{\theta}' - \boldsymbol{\theta}_0$. The distinct gradient components correspond to the cosmological sensitivity maps displayed in Fig. 5.2. The gradient is then squared and divided by the mean fiducial galaxy field and marginalized over phase realizations, as required by Eq. (5.12), to finally yield the desired Fisher information elements. The resulting Fisher information map for a particular spherical slice through the 3D field is illustrated in Fig. 5.3.

In this approach, we do not explicitly enforce that $\sum_i \Omega_i = 1$ for $\boldsymbol{\theta}'$, since we are interested in infinitesimal and independent variations of the cosmological parameters as required by the Fisher information. Therefore, although this is marginally inconsistent from the cosmological perspective, it is perfectly consistent from the information theory perspective. Furthermore, enforcing the $\sum_i \Omega_i = 1$ in our forward model introduces dependencies of the variations required for the evaluation of the Fisher information, which renders the interpretation more convoluted. Nonetheless, we also tested the impact of enforcing $\sum_i \Omega_i = 1$ in our forward model and found that this has a negligible effect on the derived Fisher information map. In addition, we verified the robustness of our results, as expected, to the choice of fiducial cosmology.

5.6 Appendix C: Marginalizing phase realizations and targeted searches

The derivation presented in the previous section took care of explicitly keeping, where needed, the conditional dependence on the specific realization of the phases of the initial density field, also referred to as the initial white noise field. The reason being that in practice, the exact phase information of the initial density field is not available:

$$\mathcal{I}(\boldsymbol{\theta}) = \int d\phi \mathcal{P}(\phi) \mathcal{I}(\boldsymbol{\theta}|\phi), \quad (5.14)$$

5.6 Appendix C: Marginalizing phase realizations and targeted searches 163

where $\mathcal{P}(\phi)$ is a Gaussian prior with zero mean and unit variance for the initial white noise field. We effectively use a Markov approximation of the integral by drawing random realizations from the white noise prior and evaluating the corresponding averaged Fisher information as:

$$\mathcal{I}(\boldsymbol{\theta}) = \frac{1}{N_\phi} \sum_i \mathcal{I}(\boldsymbol{\theta}|\phi_i). \quad (5.15)$$

where the ϕ_i are independent white noise realizations. Similarly, we can just decompose the averaged Fisher information into the individual components per volume element:

$$h_j(\boldsymbol{\theta}) = \frac{1}{N_\phi} \sum_i h_j(\boldsymbol{\theta}|\phi_i). \quad (5.16)$$

It is clear that since we marginalize over all possible white noise realizations, there cannot be any variation in the $h_i(\boldsymbol{\theta})$ since every configuration is equally likely. Thus, without any further prior information on the phases, every point in the Universe, on average, is expected to contribute exactly the same amount of information.

This is one of the logical reasons behind performing homogeneous cosmological surveys. However, we are no longer in a regime of cosmology where complete ignorance about the Universe prevails. We have access to data, which can inform us about the specific realization of our Universe. If we account for this fact, we can condition the average Fisher information on the existing data:

$$\mathcal{I}(\boldsymbol{\theta}|\mathbf{d}) = \int d\phi \mathcal{P}(\phi|\mathbf{d}) \mathcal{I}(\boldsymbol{\theta}|\phi). \quad (5.17)$$

Now, $\mathcal{P}(\phi|\mathbf{d})$ is the data constrained posterior of the initial phase distribution. Obtaining this posterior distribution is a non-trivial task, but one that is solved by our BORG algorithm. In particular, BORG provides a Markov approximation to the high-dimensional posterior distribution $\mathcal{P}(\phi|\mathbf{d})$, such that we can approximate the integral as:

$$\mathcal{I}(\boldsymbol{\theta}|\mathbf{d}) = \frac{1}{N_\phi^{\text{BORG}}} \sum_i \mathcal{I}(\boldsymbol{\theta}|\phi_i^{\text{BORG}}), \quad (5.18)$$

with this inferred posterior of initial phase distributions being fairly robust to the details of the physical model adopted [112]. Similarly, we can use the results of BORG to estimate the Fisher information per volume element as:

$$h_j(\boldsymbol{\theta}|\mathbf{d}) = \frac{1}{N_\phi^{\text{BORG}}} \sum_i h_j(\boldsymbol{\theta}|\phi_i^{\text{BORG}}). \quad (5.19)$$

The key point to realize here is that the result does not depend on the white noise phases ϕ as they have been marginalized out. More precisely, this has been achieved by evaluating Eq. (5.19) with $N_\phi^{\text{BORG}} = 2422$ samples characterizing the posterior $\mathcal{P}(\phi|\mathbf{d})$.

The large-scale structure provides a wealth of cosmological information. One of the most intriguing aspects is that it allows for obtaining insight into how the initial conditions of our own universe looked like, allowing for constraining fundamental physics from large-scale structure tracer data along the way. However, this cosmological information is hidden in the highly non-linear structures formed throughout the history of the universe. In order to extract it, one needs to employ sophisticated forward modelling techniques. The aim of this thesis was precisely to make one step closer towards achieving this goal.

The framework developed in this thesis is the field-level EFT approach to large-scale structure described in Chapter 3 and Section 2.6.2. It allows for extending the usual standard perturbation theory approach into the quasi-linear regime and allows for rigorous control of the theoretical uncertainties on the obtained cosmological constraints. Furthermore, it clearly motivates the shape of the likelihood to be utilized in the inference, and gives a clear model for the matter-tracer mapping in perturbation theory. The EFT forward model we develop, working at the field-level, is capable of modelling the final density field mode by mode and therefore extracting maximum available information from the measured tracer field. This in turn allows for reproducing leading- and next-to-leading order summary statistics of the final tracer field at the given order in perturbation theory, and allows for a systematic way of proceeding to higher order n -point functions (see Section 2.4 and Section 2.6.2).

In Chapter 3, we present a field-level forward modelling approach using the EFT of large-scale structure capable of jointly sampling the cosmology (in particular the amplitude σ_8), bias parameters as well as marginalizing over plausible initial conditions consistent with the given data realization. This work is one of the first to allow for consistent marginalization of the initial conditions space within the EFT framework. We employ our EFT models in scenarios with model-mismatch and demonstrate that

the consistency tests we present are a powerful tool of determining whether the forward model captures all relevant clustering physics at the scales considered, as well as test for the forward model robustness. We demonstrate the ability of unbiased inference of the σ_8 parameter even in the presence of strong model mismatch between the generated data and model.

Afterward, in Chapter 4, we employ the reconstructed local dark matter density field provided by the BORG algorithm, using cosmology consistent with Planck 2018 results [5] in order to constrain fundamental dark matter properties such as its decay time and annihilation cross section from the available γ -ray sky data provided by Fermi-LAT [178]. We consider different plausible decay and annihilation channels. For annihilation, we are able to exclude the WIMP thermal relic cross section for the case of s -wave for masses $m_\chi \lesssim 7 \text{ GeV}/c^2$, while for velocity-dependent channels, p - and d -wave, we obtain $a_1 < 2.4 \times 10^{-21} \text{ cm}^3 \text{ s}^{-1}$ and $a_2 < 3.0 \times 10^{-18} \text{ cm}^3 \text{ s}^{-1}$ at 95% confidence for $m_\chi = 10 \text{ GeV}/c^2$ respectively, both significantly above the corresponding thermal relic cross section as calculated in Section 4.5.5. For the WIMP decay, we detect at the level of 3.3σ a cross correlation between the measured γ -ray sky flux and the decay template we constructed from the local dark matter distribution. The inferred decay rate is $\Gamma \approx 6 \times 10^{-28} \text{ s}^{-1}$, however, the γ -ray data seems to prefer better a power-law spectrum with an index $p = -2.75_{-0.46}^{+0.71}$, suggesting astrophysical origin. With this we demonstrate the potential of field-level large-scale structure reconstructions, paving the way for similar considerations in the future, where the field-level forward modelling can be used in constraining fundamental physics.

Finally, in Chapter 5, we propose an active learning framework capable of detecting regions on the sky promising highest information gain on the cosmology of interest. In the era of big all-sky surveys, such a framework can be beneficial, allowing for focusing only on particular highest information gain regions on the sky, and thus providing faster convergence towards most optimal cosmological constraints. We construct such information maps of galaxy count data for a particular forward model implemented within the BORG algorithm, and demonstrate that the information map constructed correctly points towards infalling regions as containing the highest information gain on cosmology. Even though an expected result was constructed, the method of weighing the Fisher information map by the posterior attained from previous Bayesian inference approaches, as presented in Section 5.6, can be applied to any forward model of interest and guide future explorations of our own universe.

APPENDIX A

MOMENTS OF THE COLD DARK MATTER BOLTZMANN EQUATION

In Section 2.1 we have encountered the collisionless Boltzmann equation as a good description of the ensemble dynamics for cold dark matter. The equation was written in Eq. (2.15) and states the following

$$\frac{\partial f}{\partial t} + \frac{\partial f}{\partial x^i} \frac{p^i}{ma} - \frac{\partial f}{\partial p^i} \left(H p^i + \frac{m}{a} \frac{\partial \Psi}{\partial x^i} \right) = 0, \quad (\text{A.1})$$

with x^i and p^i being the corresponding phase space position and momentum, f being the distribution function, H the Hubble parameter, a the scale factor and Ψ the gravitational potential in which dark matter particles evolve. The goal of this appendix is to derive the fluid equations from Eq. (2.16) and subsequently show why the vanishingly small velocity dispersion of dark matter allows us to drop the stress tensor. However, this approximation breaks down once we start to discuss structure formation as explained in Section 2.3 and Section 2.4.

In order to derive the fluid equations, we need to take the first two moments of Eq. (A.1). The reason why we can stop at the level of second moment was presented at the end of Section 2.3 and is a direct consequence of the smallness of the mean free path of the dark matter particles with respect to the typical scales of interest. We will list the

first moment results for each term from Eq. (A.1), explicitly writing out the arguments

$$\begin{aligned}
\int_{\mathbf{p}} \partial_t f(\mathbf{x}, \mathbf{p}, t) &= \partial_t \int_{\mathbf{p}} f(\mathbf{x}, \mathbf{p}, t) = \partial_t n(\mathbf{x}, t) = \frac{1}{m} \partial_t \rho(\mathbf{x}, t), \\
\int_{\mathbf{p}} \partial_i f(\mathbf{x}, \mathbf{p}, t) \frac{p^i}{am} &= \frac{1}{am} \int_{\mathbf{p}} \left[\partial_i (f(\mathbf{x}, \mathbf{p}, t) p^i) - \underbrace{f(\mathbf{x}, \mathbf{p}, t) \partial_i p^i}_{=0} \right] \\
&= \frac{1}{am} \partial_i \left(\langle m \rangle_{f(\mathbf{x}, \mathbf{p}, t)} \frac{\langle p^i \rangle_{f(\mathbf{x}, \mathbf{p}, t)}}{\langle m \rangle_{f(\mathbf{x}, \mathbf{p}, t)}} \right) \\
&= \frac{1}{am} \partial_i (\rho(\mathbf{x}, t) v^i(\mathbf{x}, t)) \\
\int_{\mathbf{p}} \left(H p^i + \frac{m}{a} \partial_i \Psi \right) \partial_{p^i} f(\mathbf{x}, \mathbf{p}, t) &= H \int_{\mathbf{p}} p^i \partial_{p^i} f(\mathbf{x}, \mathbf{p}, t) + \frac{m}{a} \int_{\mathbf{p}} \partial_i \Psi(\mathbf{x}, t) \partial_{p^i} f(\mathbf{x}, \mathbf{p}, t) \\
&= H \underbrace{p^i f(\mathbf{x}, \mathbf{p}, t) \Big|_{p \rightarrow -\infty}^{p \rightarrow \infty}}_{=0} - H \int_{\mathbf{p}} \underbrace{\partial_{p^i} p^i}_{=3} f(\mathbf{x}, \mathbf{p}, t) \\
&\quad + \frac{m}{a} \partial_i \Psi(\mathbf{x}, t) \underbrace{f(\mathbf{x}, \mathbf{p}, t) \Big|_{p \rightarrow -\infty}^{p \rightarrow \infty}}_{=0} - \frac{m}{a} \int_{\mathbf{p}} \underbrace{(\partial_{p^i} \partial_i \Psi(\mathbf{x}, t))}_{=0} f(\mathbf{x}, \mathbf{p}, t) \\
&= -\frac{1}{m} 3H \rho(\mathbf{x}, t). \tag{A.2}
\end{aligned}$$

We have relied throughout on the phase space coordinates \mathbf{x} and \mathbf{p} being independent, as well as that the distribution function has an exponentially decaying wings and approaches to zero for infinitely big momenta. In summary, the first moment of the Boltzmann equation gives

$$\partial_t \rho(\mathbf{x}, t) + \frac{1}{am} \partial_i (\rho(\mathbf{x}, t) v^i(\mathbf{x}, t)) + 3H \rho(\mathbf{x}, t) = 0, \tag{A.3}$$

which is exactly the continuity equation. Now, using the relation $\rho(\mathbf{x}, t) = \rho(t)(1 + \delta(\mathbf{x}, t))$ to introduce density fluctuations $\delta(\mathbf{x}, t)$, alongside with the energy conservation equation Eq. (2.9) to remove the $\sim H\rho(\mathbf{x}, t)$ term, immediately gives the first line in Eq. (2.17).

In order to derive the Euler equation, we can take the second moment of Eq. (A.1).

Again, term by term we get

$$\begin{aligned}
\partial_t \int_{\mathbf{p}} p^i f(\mathbf{x}, \mathbf{p}, t) &= \partial_t (\rho(\mathbf{x}, t) v^i(\mathbf{x}, t)) \\
\int_{\mathbf{p}} p^j \partial_i f(\mathbf{x}, \mathbf{p}, t) \frac{p^i}{am} &= \frac{1}{am} \partial_i \int_{\mathbf{p}} f(\mathbf{x}, \mathbf{p}, t) p^i p^j \\
&\stackrel{\text{def.}}{=} \frac{1}{a} (\rho(\mathbf{x}, t) v^i(\mathbf{x}, t) v^j(\mathbf{x}, t) + \sigma^{ij}(\mathbf{x}, t)) \\
\int_{\mathbf{p}} p^j \left(H p^i + \frac{m}{a} \partial_i \Psi(\mathbf{x}, t) \right) \partial_{p^i} f(\mathbf{x}, \mathbf{p}, t) &\stackrel{\text{P.I.}}{=} -H \int_{\mathbf{p}} \partial_{p^i} (p^i p^j) f(\mathbf{x}, \mathbf{p}, t) \\
&\quad - \frac{m}{a} \int_{\mathbf{p}} \partial_{p^i} (p^j \partial_i \Psi(\mathbf{x}, t)) f(\mathbf{x}, \mathbf{p}, t) \\
&= -H \int_{\mathbf{p}} (p^i \delta_{ij} + 3p^j) f(\mathbf{x}, \mathbf{p}, t) \\
&\quad - \frac{m}{a} \int_{\mathbf{p}} \delta_{ij} \partial_i \Psi(\mathbf{x}, t) f(\mathbf{x}, \mathbf{p}, t) \\
&= -4H \rho(\mathbf{x}, t) v^j(\mathbf{x}, t) - \frac{1}{a} \partial_j \Psi(\mathbf{x}, t) \rho(\mathbf{x}, t),
\end{aligned} \tag{A.4}$$

where $\stackrel{\text{P.I.}}{=}$ stands for partial integration. Once again, we have used the asymptotic properties of the distribution function to neglect other terms appearing after partial integration. We also introduced the anisotropic stress tensor $\sigma^{ij}(\mathbf{x}, t)$, which quantifies departures from coherent flows. Reader is referred to [321] for classical treatment and [322] for relativistic treatment of the anisotropic stress tensor. In total, the 2nd moment of the Boltzmann equation gives

$$\begin{aligned}
\partial_t (\rho(\mathbf{x}, t) v^i(\mathbf{x}, t)) + \frac{1}{a} \partial_i (\rho(\mathbf{x}, t) v^i(\mathbf{x}, t) v^j(\mathbf{x}, t) + \sigma^{ij}(\mathbf{x}, t)) \\
+ H \rho(\mathbf{x}, t) v^j(\mathbf{x}, t) + \frac{1}{a} \partial_j \Psi(\mathbf{x}, t) \rho(\mathbf{x}, t) = 0.
\end{aligned} \tag{A.5}$$

After accounting for the appearance of continuity equation in the above, it reduces to the familiar Euler equation, given in the second line of Eq. (2.17)

$$v^j \prime(\mathbf{x}, t) + v^i(\mathbf{x}, t) \partial_i v^j(\mathbf{x}, t) + \mathcal{H} v^j(\mathbf{x}, t) + \partial_j \Psi(\mathbf{x}, t). \tag{A.6}$$

We have also dropped the anisotropic stress from the final equation, which however should be taken into account once discussing structure formation (see Section 2.3). The absence of anisotropic tensor outside the regime of structure formation, follows from the assumption that dark matter is dynamically cold. This means that the dark matter

particles are moving along with the bulk flow, and the velocity field has negligible velocity dispersion. In other words, the momentum transfer between the dark matter particles is assumed to be completely reversible, and comes from the transport of the dark matter particles due to the surrounding gravitational potential. This means that in principle one can write down the distribution function of dark matter as

$$f(\mathbf{x}, \mathbf{p}, t) \sim n(\mathbf{x}, t) \delta_D^{(3)}(\mathbf{p} - m\mathbf{v}(\mathbf{x}, t)). \quad (\text{A.7})$$

To see that this distribution function has vanishing anisotropic stress, it is enough to again calculate $\langle p^i p^j \rangle$ and see it only produces the first term on the third line of Eq. (A.4). Once structure formation ensues, this approximation no longer holds, since the fluid flow lines begin intersecting.

APPENDIX B

LAGRANGIAN PERTURBATION THEORY

In Chapter 2 we have derived how the solutions for the density contrast and curl-free velocity field look like in Eulerian perturbation theory. However, given that we are in pursuit of describing motions of the dark matter fluid, a more natural frame to treat the particle trajectories is the Lagrangian frame. One of the main advantages of this approach is faster convergence towards the true clustering solution given by direct N-body integration, for example. In Lagrangian PT it is possible to also describe much larger density contrasts compared to Eulerian PT, which relies on the fact that density contrast remains small as emphasized at the end of Section 2.1. Furthermore, in one dimensional universe, the first order LPT solutions to the clustering problem are exact up to the crossing of fluid flow trajectories [323, 324]. These properties also motivate our own EFT forward modelling choices presented in Chapter 3.

As in Eulerian PT, the LPT seeks to find solutions to the dynamics of the self-gravitating system such as the dark matter fluid. Again, as emphasized in Section 2.1, we are not interested in following each individual particle, but want to describe the evolution of the motion of an ensemble of dark matter particles, or equivalently, the fluid elements. Numerically, this can be achieved by discretizing the phase-space sheet of dark matter and integrating the geodesic equation of dark matter particles. This is exactly the approach undertaken by numerical N-body simulations (see e.g. [325]). The geodesic equations for non-relativistic dark matter are (see e.g. Section 3.3.2 in [42])

$$\begin{aligned}\frac{dx^i}{d\eta} &= \frac{p^i}{m} \\ \frac{dp^i}{d\eta} &= -\mathcal{H}p^i - m\partial_i\Psi,\end{aligned}\tag{B.1}$$

using the same notation as in Section 2.1. The LPT proceeds by introducing the displacement field $\boldsymbol{\psi}$ relating the initial Lagrangian coordinates, \mathbf{q} , i.e. the initial point along the given fluid flow line, and final Eulerian coordinates \mathbf{x} by

$$\mathbf{x}(\mathbf{q}, t) = \mathbf{q} + \boldsymbol{\psi}(\mathbf{q}, t).\tag{B.2}$$

The corresponding Jacobian of this coordinate transform is

$$\mathcal{J} = \left| \frac{d^3 \mathbf{x}}{d^3 \mathbf{q}} \right| = \det(\mathbb{1} + \partial_{\mathbf{q}} \boldsymbol{\psi}), \quad (\text{B.3})$$

which then, due to mass conservation, imposes

$$1 + \delta(\mathbf{x}) = \mathcal{J}^{-1} \quad (\text{B.4})$$

assuming initial slice fluctuations are very small. Combining then the two first order ODEs from Eq. (B.1) into a single 2nd order ODE for \mathbf{x} and focusing on the divergence part of the equation, we get

$$\nabla_{\mathbf{x}} \left[\frac{d^2 \mathbf{x}}{d^2 \eta} + \mathcal{H} \frac{d \mathbf{x}}{d \eta} \right] = \frac{3}{2} \Omega_m \mathcal{H}^2 (\mathcal{J}^{-1} - 1), \quad (\text{B.5})$$

assuming EdS as in Section 2.1. Note that the neglect of the curl part of the equation of motion only affects us at third order (see e.g. [30]). Furthermore, the error in taking the EdS approximation with respect to the full Λ CDM treatment is below a percent (*ibid.*). Since in Chapter 3 we don't go further than 2nd order LPT and therefore there is no need to discuss the higher order solutions in this review of LPT. However, for a full systematic treatment, the reader is referred to [30, 116, 297, 326]. Note that the derivatives with respect to Eulerian and Lagrangian coordinates are related as $\partial_{x^i} = \mathcal{J}^{-1} \nabla_{q^i}$. One can now continue and expand Eq. (B.5) in displacement field perturbations. We need to expand both \mathcal{J} and $\boldsymbol{\psi}$, keeping all relevant terms at the given order. At first order we have

$$\mathcal{J}^{(1)} = \sum_i \partial_{q^i} \psi_i^{(1)} \equiv \sum_i \psi_{i,i}^{(1)}. \quad (\text{B.6})$$

This then results in the following differential equation for the displacement field $\boldsymbol{\psi}$ (see also [54])

$$\left(\frac{d^2}{d\eta^2} + \mathcal{H} \frac{d}{d\eta} - \frac{3}{2} \Omega_m \mathcal{H}^2 \right) \psi_{i,i}^{(1)} = 0, \quad (\text{B.7})$$

which is the same as the linear growth equation from Eq. (2.21). Therefore, we once again have the decaying and growing solution. Focusing on the growing solution and using the relation between the displacement field and density field at leading order from Eq. (B.4) we have

$$\boldsymbol{\psi}^{(1)}(\mathbf{k}, \eta) = \frac{i\mathbf{k}}{k^2} \delta^{(1)}(\mathbf{k}, \eta). \quad (\text{B.8})$$

This solution is also known as the Zel'dovich approximation [323] and can be demonstrated to yield surprisingly accurate results with respect to the full N-body solution (see for example Fig 6. in [297]). In fact, in one dimensional universe this displacement field yields exact solutions when applied in Eq. (B.2) to displace particles from initial

positions \mathbf{q} , however only up to the shell crossing. Remedying this problem requires the introduction of counter terms via the stress tensor, τ^{ij} , as discussed in Section 2.3.

This approximation can be improved quite significantly when proceeding to second order solution. For this, one can again use the fact that in EdS the growth factors are simple and at n^{th} order are given as n powers of a , hence the following relations hold

$$\frac{d^2\boldsymbol{\psi}^{(n)}}{da^2} = \frac{n(n-1)}{a^2}\boldsymbol{\psi}^{(n)}, \quad \frac{d\boldsymbol{\psi}^{(n)}}{da} = n\boldsymbol{\psi}^{(n)}. \quad (\text{B.9})$$

Therefore at n^{th} order, Eq. (B.5) is given by [54]

$$\mathcal{J}(\mathbb{1} + \partial_{\mathbf{q}}\boldsymbol{\psi})^{-1} \left[n^2 + \frac{n}{2} \right] \mathcal{H}^2 \psi_{i,j}^{(n)} = \frac{3}{2} \mathcal{H}^2 (\mathcal{J} - 1). \quad (\text{B.10})$$

Note that we have kept the Jacobian inverse, since we need to expand all the operators up to 2nd order explicitly before solving the differential equation. Expanding now all the Jacobian terms and keeping everything up to 2nd order in displacement fields gives the following equation for $\boldsymbol{\psi}^{(2)}$

$$\psi_{i,i}^{(2)} = -\frac{3}{14} \sum_{i,j} \left(\psi_{i,i}^{(1)} \psi_{j,j}^{(1)} - \psi_{i,j}^{(1)} \psi_{j,i}^{(1)} \right), \quad (\text{B.11})$$

and the resulting 2nd order displacement field $\boldsymbol{\psi}^{(2)}$ can be obtained by direct integration. In fact, in Appendix A of [30] analytical solutions up to any order are given for the components of the displacement field in the EdS approximation.

After calculating the displacement field up to any given order, one can then directly calculate the final density field fluctuations which arises as a result through

$$\delta(\mathbf{k}) = \int d^3x e^{-i\mathbf{k}\cdot\mathbf{x}} \delta(\mathbf{x}) = \int d^3q e^{-i\mathbf{k}\cdot\mathbf{q}} (e^{-i\mathbf{k}\cdot\boldsymbol{\psi}(\mathbf{q})} - 1), \quad (\text{B.12})$$

using the Jacobian from Eq. (B.3). Proceeding and expanding the exponential of the above equation gives order by order the relation between Eulerian and Lagrangian PT (see for example [297]). This relation is only achievable, however, before shell crossing, i.e. only while the Jacobian \mathcal{J} is invertible and non-singular (see [327] for a proof).

An important note is that instead of expanding in density fluctuations, what LPT does is to expand in small displacements. It doesn't assume anything about the final density fluctuations, and hence can capture a much wider range of final overdensities. The downturn is that it can only be taken as formally correct up until the fluid trajectories cross. Nonetheless, LPT does provide insights into the fully non-linear regime and is capable of reproducing leading moments of the fully non-linear density field quite accurately (see [30, 297, 328]). Precisely because of this reason, we employ LPT in our own EFT forward models in Chapter 3.

BIBLIOGRAPHY

- [1] G. F. Smoot, C. L. Bennett, A. Kogut, et al., *Structure in the COBE Differential Microwave Radiometer First-Year Maps*, ApJ **396** (Sept., 1992) L1.
- [2] E. Komatsu, K. M. Smith, J. Dunkley, et al., *Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation*, ApJS **192** (Feb., 2011) [[arXiv:1001.4538](#)].
- [3] G. Hinshaw, D. Larson, E. Komatsu, et al., *Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Parameter Results*, ApJS **208** (Oct., 2013) 19, [[arXiv:1212.5226](#)].
- [4] Planck Collaboration, *Planck 2013 results. XVI. Cosmological parameters*, A&A **571** (Nov, 2014) A16, [[arXiv:1303.5076](#)].
- [5] Planck Collaboration, N. Aghanim, Y. Akrami, et al., *Planck 2018 results. VI. Cosmological parameters*, A&A **641** (Sept., 2020) A6, [[arXiv:1807.06209](#)].
- [6] DESI Collaboration, *The DESI Experiment Part I: Science, Targeting, and Survey Design*, arXiv e-prints (Oct., 2016) [arXiv:1611.00036](#), [[arXiv:1611.00036](#)].
- [7] L. Amendola, S. Appleby, A. Avgoustidis, et al., *Cosmology and fundamental physics with the Euclid satellite*, Living Reviews in Relativity **21** (Apr., 2018) 2, [[arXiv:1606.00180](#)].
- [8] M. Takada, R. S. Ellis, M. Chiba, et al., *Extragalactic science, cosmology, and galactic archaeology with the Subaru prime focus spectrograph*, Publications of the Astronomical Society of Japan **66** (2014), no. 1.
- [9] O. Doré, J. Bock, M. Ashby, et al., *Cosmology with the SPHEREX All-Sky Spectral Survey*, arXiv e-prints (Dec., 2014) [arXiv:1412.4872](#), [[arXiv:1412.4872](#)].

- [10] P. A. Abell, J. Allison, S. F. Anderson, et al., *Lsst science book, version 2.0*, arXiv preprint arXiv:0912.0201 (2009).
- [11] S. Alam, M. Aubert, S. Avila, et al., *Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Cosmological implications from two decades of spectroscopic surveys at the Apache Point Observatory*, *Phys. Rev. D* **103** (Apr., 2021) 083533, [arXiv:2007.08991].
- [12] H. Gil-Marín, J. Noreña, L. Verde, et al., *The power spectrum and bispectrum of SDSS DR11 BOSS galaxies - I. Bias and gravity*, *MNRAS* **451** (July, 2015) 539–580, [arXiv:1407.5668].
- [13] H. Gil-Marín, W. J. Percival, L. Verde, et al., *The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: RSD measurement from the power spectrum and bispectrum of the DR12 BOSS galaxies*, *MNRAS* **465** (Feb., 2017) 1757–1788, [arXiv:1606.00439].
- [14] O. H. E. Philcox and M. M. Ivanov, *BOSS DR12 full-shape cosmology: Λ CDM constraints from the large-scale galaxy power spectrum and bispectrum monopole*, *Phys. Rev. D* **105** (Feb., 2022) 043517, [arXiv:2112.04515].
- [15] G. D’Amico, Y. Donath, M. Lewandowski, L. Senatore, and P. Zhang, *The BOSS bispectrum analysis at one loop from the Effective Field Theory of Large-Scale Structure*, arXiv e-prints (June, 2022) arXiv:2206.08327, [arXiv:2206.08327].
- [16] D. Gualdi and L. Verde, *Integrated trispectrum detection from BOSS DR12 NGC CMASS*, *J. Cosmology Astropart. Phys.* **2022** (Sept., 2022) 050, [arXiv:2201.06932].
- [17] O. H. E. Philcox, *Probing parity violation with the four-point correlation function of BOSS galaxies*, *Phys. Rev. D* **106** (Sept., 2022) 063501, [arXiv:2206.04227].
- [18] J. Hou, Z. Slepian, and R. N. Cahn, *Measurement of Parity-Odd Modes in the Large-Scale 4-Point Correlation Function of SDSS BOSS DR12 CMASS and LOWZ Galaxies*, arXiv e-prints (June, 2022) arXiv:2206.03625, [arXiv:2206.03625].
- [19] J. Jasche and B. D. Wandelt, *Bayesian inference from photometric redshift surveys*, *MNRAS* **425** (Sept., 2012) 1042–1056, [arXiv:1106.2757].
- [20] J. Jasche and B. D. Wandelt, *Bayesian physical reconstruction of initial conditions from large-scale structure surveys*, *MNRAS* **432** (June, 2013) 894–913, [arXiv:1203.3639].

- [21] J. Jasche, F. S. Kitaura, B. D. Wandelt, and T. A. Enßlin, *Bayesian power-spectrum inference for large-scale structure data*, MNRAS **406** (July, 2010) 60–85, [arXiv:0911.2493].
- [22] J. Jasche, F. Leclercq, and B. D. Wandelt, *Past and present cosmic structure in the SDSS DR7 main sample*, J. Cosmology Astropart. Phys. **2015** (Jan., 2015) 036, [arXiv:1409.6308].
- [23] J. Jasche and G. Lavaux, *Physical Bayesian modelling of the non-linear matter distribution: New insights into the nearby universe*, A&A **625** (May, 2019) A64, [arXiv:1806.11117].
- [24] M. C. Neyrinck, *ZOBOV: a parameter-free void-finding algorithm*, MNRAS **386** (June, 2008) 2101–2109, [arXiv:0712.3049].
- [25] D. Kodi Ramanah, T. Charnock, and G. Lavaux, *Painting halos from cosmic density fields of dark matter with physically motivated neural networks*, Phys. Rev. D **100** (Aug., 2019) 043515, [arXiv:1903.10524].
- [26] T. Charnock, G. Lavaux, B. D. Wandelt, et al., *Neural physical engines for inferring the halo mass distribution function*, MNRAS **494** (May, 2020) 50–61, [arXiv:1909.06379].
- [27] D. Baumann, A. Nicolis, L. Senatore, and M. Zaldarriaga, *Cosmological non-linearities as an effective fluid*, J. Cosmology Astropart. Phys. **7** (July, 2012) 51, [arXiv:1004.2488].
- [28] S. M. Carroll, S. Leichenauer, and J. Pollack, *Consistent effective theory of long-wavelength cosmological perturbations*, Phys. Rev. D **90** (July, 2014) 023518, [arXiv:1310.2920].
- [29] J. J. M. Carrasco, M. P. Hertzberg, and L. Senatore, *The effective field theory of cosmological large scale structures*, Journal of High Energy Physics **2012** (Sept., 2012) 82, [arXiv:1206.2926].
- [30] F. Schmidt, *An n -th order Lagrangian forward model for large-scale structure*, J. Cosmology Astropart. Phys. **2021** (Apr., 2021) 033, [arXiv:2012.09837].
- [31] V. Desjacques, D. Jeong, and F. Schmidt, *Large-scale galaxy bias*, Phys. Rep. **733** (Feb., 2018) 1–193, [arXiv:1611.09787].
- [32] S. Mukherjee and B. D. Wandelt, *Making maps of cosmological parameters*, J. Cosmology Astropart. Phys. **2018** (Jan., 2018) [arXiv:1712.01986].

- [33] B. A. Bassett, *Optimizing cosmological surveys in a crowded market*, Physical Review D **71** (2005), no. 8.
- [34] T. A. Enßlin, M. Frommert, and F. S. Kitaura, *Information field theory for cosmological perturbation reconstruction and nonlinear signal analysis*, Phys. Rev. D **80** (Nov., 2009) 105005, [[arXiv:0806.3474](#)].
- [35] T. A. Enßlin and M. Frommert, *Reconstruction of signals with unknown spectra in information field theory with parameter uncertainty*, Phys. Rev. D **83** (May, 2011) 105014, [[arXiv:1002.2928](#)].
- [36] M. Selig, M. R. Bell, H. Junklewitz, et al., *NIFTY - Numerical Information Field Theory. A versatile PYTHON library for signal inference*, A&A **554** (June, 2013) A26, [[arXiv:1301.4499](#)].
- [37] M. Guardiani, P. Frank, A. Kostić, et al., *Causal, bayesian, & non-parametric modeling of the sars-cov-2 viral load distribution vs. patient's age*, Plos one **17** (2022), no. 10 e0275011.
- [38] M. Guardiani, P. Frank, A. Kostić, and T. Enßlin, *Towards moment-constrained causal modeling*, in Physical Sciences Forum, vol. 5, p. 7, MDPI, 2022.
- [39] V. F. Mukhanov and G. V. Chibisov, *Quantum fluctuations and a nonsingular universe*, ZhETF Pisma Redaktsiiu **33** (May, 1981) 549–553.
- [40] V. F. Mukhanov, V. Mukhanov, and M. Viatcheslav, Physical foundations of cosmology. Cambridge university press, 2005.
- [41] S. Weinberg, Cosmology. OUP Oxford, 2008.
- [42] S. Dodelson and F. Schmidt, Modern cosmology. Academic press, 2020.
- [43] S. Weinberg, Gravitation and cosmology: principles and applications of the general theory of relativity. John Wiley & Sons, 1972.
- [44] A. Friedmann, *Über die Krümmung des Raumes*, Zeitschrift für Physik **10** (1922) 377–386.
- [45] A. Friedmann, *Über die Möglichkeit einer Welt mit konstanter negativer Krümmung des Raumes*, Zeitschrift für Physik **21** (Dec., 1924) 326–332.
- [46] G. Lemaître, *Expansion of the universe, A homogeneous universe of constant mass and increasing radius accounting for the radial velocity of extra-galactic nebulae*, MNRAS **91** (Mar., 1931) 483–490.

- [47] H. P. Robertson, *Kinematics and World-Structure*, ApJ **82** (Nov., 1935) 284.
- [48] H. P. Robertson, *Kinematics and World-Structure II.*, ApJ **83** (Apr., 1936) 187.
- [49] H. P. Robertson, *Kinematics and World-Structure III.*, ApJ **83** (May, 1936) 257.
- [50] A. G. Walker, *On Milne's theory of world-structure*, Proceedings of the London Mathematical Society **42** (June, 1937) 90.
- [51] E. Bertschinger, *Cosmological dynamics*, arXiv preprint astro-ph/9503125 (1995).
- [52] C.-P. Ma and E. Bertschinger, *Cosmological Perturbation Theory in the Synchronous and Conformal Newtonian Gauges*, ApJ **455** (Dec., 1995) 7, [[arXiv:astro-ph/astro-ph/9506072](#)].
- [53] S. Hutschenreuter, S. Dorn, J. Jasche, et al., *The primordial magnetic field in our cosmic backyard*, Classical and Quantum Gravity **35** (Aug., 2018) [[arXiv:1803.02629](#)].
- [54] T. Baldauf, *Effective Field Theory of Large-Scale Structure: Lecture Notes of the Les Houches Summer School*, Oxford University Press, 2020.
- [55] F. Bernardeau, S. Colombi, E. Gaztañaga, and R. Scoccimarro, *Large-scale structure of the Universe and cosmological perturbation theory*, Phys. Rep. **367** (Sept., 2002) 1–248, [[astro-ph/0](#)].
- [56] B. W. Lee and S. Weinberg, *Cosmological lower bound on heavy-neutrino masses*, Phys. Rev. Lett. **39** (July, 1977) 165–168.
- [57] G. Steigman and M. S. Turner, *Cosmological constraints on the properties of weakly interacting massive particles*, Nuclear Physics B **253** (Jan., 1985) 375–386.
- [58] J. E. Gunn, B. W. Lee, I. Lerche, D. N. Schramm, and G. Steigman, *Some astrophysical consequences of the existence of a heavy stable neutral lepton.*, ApJ **223** (Aug., 1978) 1015–1031.
- [59] G. Jungman, M. Kamionkowski, and K. Griest, *Supersymmetric dark matter*, Phys. Rep. **267** (Mar., 1996) 195–373, [[arXiv:hep-ph/hep-ph/9506380](#)].
- [60] G. Steigman, B. Dasgupta, and J. F. Beacom, *Precise relic WIMP abundance and its impact on searches for dark matter annihilation*, Phys. Rev. D **86** (July, 2012) [[arXiv:1204.3622](#)].
- [61] M. Laine and Y. Schröder, *Quark mass thresholds in QCD thermodynamics*, Phys. Rev. D **73** (Apr., 2006) [[arXiv:hep-ph/hep-ph/0603048](#)].

- [62] D. Jeong and E. Komatsu, *Perturbation Theory Reloaded: Analytical Calculation of Nonlinearity in Baryonic Oscillations in the Real-Space Matter Power Spectrum*, ApJ **651** (Nov., 2006) 619–626, [arXiv:astro-ph/astro-ph/0604075].
- [63] D. Jeong and E. Komatsu, *Perturbation Theory Reloaded. II. Nonlinear Bias, Baryon Acoustic Oscillations, and Millennium Simulation in Real Space*, ApJ **691** (Jan., 2009) 569–595, [arXiv:0805.2632].
- [64] A. A. Abolhasani, M. Mirbabayi, and E. Pajer, *Systematic renormalization of the effective theory of large scale structure*, Journal of Cosmology and Astroparticle Physics **2016** (2016).
- [65] B. Jain and E. Bertschinger, *Second-Order Power Spectrum and Nonlinear Evolution at High Redshift*, ApJ **431** (Aug., 1994) 495, [arXiv:astro-ph/astro-ph/9311070].
- [66] M. Mirbabayi, F. Schmidt, and M. Zaldarriaga, *Biased tracers and time evolution*, J. Cosmology Astropart. Phys. **7** (July, 2015) 30, [arXiv:1412.5169].
- [67] T. Baldauf, L. Mercolli, M. Mirbabayi, and E. Pajer, *The bispectrum in the Effective Field Theory of Large Scale Structure*, J. Cosmology Astropart. Phys. **2015** (May, 2015) [arXiv:1406.4135].
- [68] M. Dijkstra, *Ly- α Emitting Galaxies as a Probe of Reionisation*, PASA **31** (Oct., 2014) [arXiv:1406.7292].
- [69] R. H. Becker, X. Fan, R. L. White, et al., *Evidence for Reionization at $z \sim 6$: Detection of a Gunn-Peterson Trough in a $z=6.28$ Quasar*, AJ **122** (Dec., 2001) 2850–2857, [arXiv:astro-ph/astro-ph/0108097].
- [70] N. Aghanim, S. Majumdar, and J. Silk, *Secondary anisotropies of the CMB*, Reports on Progress in Physics **71** (June, 2008) [arXiv:0711.0518].
- [71] J. E. Carlstrom, G. P. Holder, and E. D. Reese, *Cosmology with the Sunyaev-Zel’dovich Effect*, ARA&A **40** (Jan., 2002) 643–680, [arXiv:astro-ph/astro-ph/0208192].
- [72] N.-M. Nguyen, J. Jasche, G. Lavaux, and F. Schmidt, *Taking measurements of the kinematic Sunyaev-Zel’dovich effect forward: including uncertainties from velocity reconstruction with forward modeling*, J. Cosmology Astropart. Phys. **2020** (Dec., 2020) [arXiv:2007.13721].
- [73] M. F. Morales and J. S. B. Wyithe, *Reionization and Cosmology with 21-cm Fluctuations*, ARA&A **48** (Sept., 2010) 127–171, [arXiv:0910.3010].

- [74] J. L. Bernal and E. D. Kovetz, *Line-intensity mapping: theory review with a focus on star-formation lines*, A&A Rev. **30** (Dec., 2022) [[arXiv:2206.15377](#)].
- [75] N. Kaiser, *On the spatial correlations of Abell clusters*, ApJ **284** (Sept., 1984).
- [76] J. N. Fry and E. Gaztanaga, *Biasing and hierarchical statistics in large-scale structure*, ApJ **413** (Aug., 1993) 447–452, [[astro-ph/9](#)].
- [77] P. McDonald, *Clustering of dark matter tracers: Renormalizing the bias parameters*, Phys. Rev. D **74** (Nov., 2006) [[astro-ph/0](#)].
- [78] V. Assassi, D. Baumann, D. Green, and M. Zaldarriaga, *Renormalized halo bias*, J. Cosmology Astropart. Phys. **8** (Aug., 2014) [[arXiv:1402.5916](#)].
- [79] S. Saito, T. Baldauf, Z. Vlah, et al., *Understanding higher-order nonlocal halo bias at large scales by combining the power spectrum with the bispectrum*, Phys. Rev. D **90** (Dec., 2014) [[arXiv:1405.1447](#)].
- [80] R. Angulo, M. Fasiello, L. Senatore, and Z. Vlah, *On the Statistics of Biased Tracers in the Effective Field Theory of Large Scale Structures*, [arXiv:1503.08826](#).
- [81] S. Alam, F. D. Albareti, C. Allende Prieto, et al., *The Eleventh and Twelfth Data Releases of the Sloan Digital Sky Survey: Final Data from SDSS-III*, ApJS **219** (July, 2015) [[arXiv:1501.00963](#)].
- [82] R. T. Cox, *Probability, Frequency and Reasonable Expectation*, American Journal of Physics **14** (Jan., 1946) 1–13.
- [83] E. T. Jaynes, Probability theory: the logic of science. Washington University St. Louis, MO, 1996.
- [84] E. T. Jaynes, *Information theory and statistical mechanics*, Physical review **106** (1957).
- [85] H. Jeffreys, *An Invariant Form for the Prior Probability in Estimation Problems*, Proceedings of the Royal Society of London Series A **186** (Sept., 1946) 453–461.
- [86] D. J. C. Mackay, Information Theory, Inference and Learning Algorithms. Cambridge university press, 2003.
- [87] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman PASP **125** (Mar., 2013) 306.
- [88] R. Neal, *MCMC Using Hamiltonian Dynamics*, in Handbook of Markov Chain Monte Carlo, pp. 113–162. Chapman and Hall/CRC, 2011.

- [89] M. Betancourt, *A Conceptual Introduction to Hamiltonian Monte Carlo*, arXiv e-prints (Jan., 2017) [[arXiv:1701.02434](#)].
- [90] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, Numerical recipes 3rd edition: The art of scientific computing. Cambridge university press, 2007.
- [91] R. M. Neal, *Slice Sampling*, arXiv e-prints (Sept., 2000) physics/0009028, [[arXiv:physics.data-an/physics/0009028](#)].
- [92] H. Wang, H. J. Mo, X. Yang, Y. P. Jing, and W. P. Lin, *ELUCID—Exploring the Local Universe with the Reconstructed Initial Density Field. I. Hamiltonian Markov Chain Monte Carlo Method with Particle Mesh Dynamics*, ApJ **794** (Oct., 2014) 94, [[arXiv:1407.3451](#)].
- [93] G. Lavaux, J. Jasche, and F. Leclercq, *Systematic-free inference of the cosmic matter density field from SDSS3-BOSS data*, arXiv e-prints (Sept., 2019) [arXiv:1909.06396](#), [[arXiv:1909.06396](#)].
- [94] F.-S. Kitaura, M. Ata, S. A. Rodríguez-Torres, et al., *COSMIC BIRTH: efficient Bayesian inference of the evolving cosmic web from galaxy surveys*, MNRAS **502** (Apr., 2021) 3456–3475, [[arXiv:1911.00284](#)].
- [95] C. Modi, Y. Feng, and U. Seljak, *Cosmological reconstruction from galaxy light: neural network based light-matter connection*, J. Cosmology Astropart. Phys. **2018** (Oct., 2018) [[arXiv:1805.02247](#)].
- [96] J. Jasche, F. S. Kitaura, C. Li, and T. A. Enßlin, *Bayesian non-linear large-scale structure inference of the Sloan Digital Sky Survey Data Release 7*, MNRAS **409** (Nov., 2010) 355–370, [[arXiv:0911.2498](#)].
- [97] M. Schmittfull, M. Simonović, M. M. Ivanov, O. H. Philcox, and M. Zaldarriaga, *Modeling galaxies in redshift space at the field level*, Journal of Cosmology and Astroparticle Physics **2021** (2021).
- [98] J. Stadler, F. Schmidt, and M. Reinecke, *Cosmology inference at the field level from biased tracers in redshift-space*, arXiv e-prints (Mar., 2023) [[arXiv:2303.09876](#)].
- [99] M. C. Neyrinck, M. A. Aragón-Calvo, D. Jeong, and X. Wang, *A halo bias function measured deeply into voids without stochasticity*, MNRAS **441** (June, 2014) 646–655, [[arXiv:1309.6641](#)].
- [100] P. Coles and B. Jones, *A lognormal model for the cosmological mass distribution*, MNRAS **248** (Jan., 1991) 1–13.

- [101] J. Carron and M. C. Neyrinck, *On the Inadequacy of N -point Correlation Functions to Describe Nonlinear Cosmological Fields: Explicit Examples and Connection to Simulations*, ApJ **750** (May, 2012) 28, [arXiv:1201.1444].
- [102] F. Schmidt, F. Elsner, J. Jasche, N. M. Nguyen, and G. Lavaux, *A rigorous EFT-based forward model for large-scale structure*, J. Cosmology Astropart. Phys. **2019** (Jan., 2019) 042, [arXiv:1808.02002].
- [103] I. Babić, F. Schmidt, and B. Tucci, *BAO scale inference from biased tracers using the EFT likelihood*, J. Cosmology Astropart. Phys. **2022** (Aug., 2022) 007, [arXiv:2203.06177].
- [104] F. Elsner, F. Schmidt, J. Jasche, G. Lavaux, and N.-M. Nguyen, *Cosmology inference from a biased density field using the EFT-based likelihood*, J. Cosmology Astropart. Phys. **2020** (Jan., 2020) 029, [arXiv:1906.07143].
- [105] G. Cabass and F. Schmidt, *The likelihood for LSS: stochasticity of bias coefficients at all orders*, J. Cosmology Astropart. Phys. **2020** (July, 2020) 051, [arXiv:2004.00617].
- [106] F. Schmidt, G. Cabass, J. Jasche, and G. Lavaux, *Unbiased cosmology inference from biased tracers using the EFT likelihood*, J. Cosmology Astropart. Phys. **2020** (Nov., 2020) 008, [arXiv:2004.06707].
- [107] F. Schmidt, *Sigma-eight at the percent level: the EFT likelihood in real space*, J. Cosmology Astropart. Phys. **2021** (Apr., 2021) 032, [arXiv:2009.14176].
- [108] M. Schmittfull, M. Simonović, V. Assassi, and M. Zaldarriaga, *Modeling biased tracers at the field level*, Phys. Rev. D **100** (Aug., 2019) 043514, [arXiv:1811.10640].
- [109] A. Kostić, N.-M. Nguyen, F. Schmidt, and M. Reinecke, *Consistency tests of field level inference with the EFT likelihood*, arXiv e-prints (Dec., 2022) [arXiv:2212.07875].
- [110] G. Cabass and F. Schmidt, *The EFT likelihood for large-scale structure*, J. Cosmology Astropart. Phys. **2020** (Apr., 2020) 042, [arXiv:1909.04022].
- [111] N. Porqueres, D. Kodi Ramanah, J. Jasche, and G. Lavaux, *Explicit Bayesian treatment of unknown foreground contaminations in galaxy surveys*, A&A **624** (Apr., 2019) A115, [arXiv:1812.05113].
- [112] N.-M. Nguyen, F. Schmidt, G. Lavaux, and J. Jasche, *Impacts of the physical data model on the forward inference of initial conditions from biased tracers*, J. Cosmology Astropart. Phys. **2021** (Mar., 2021) 058, [arXiv:2011.06587].

- [113] O. H. E. Philcox, M. M. Ivanov, G. Cabass, et al., *Cosmology with the redshift-space galaxy bispectrum monopole at one-loop order*, Phys. Rev. D **106** (Aug., 2022) [[arXiv:2206.02800](#)].
- [114] C. Rampf, *The recursion relation in Lagrangian perturbation theory*, J. Cosmology Astropart. Phys. **12** (Dec., 2012) 4, [[arXiv:1205.5274](#)].
- [115] V. Zheligovsky and U. Frisch, *Time-analyticity of Lagrangian particle trajectories in ideal fluid flow*, Journal of Fluid Mechanics **749** (June, 2014) 404–430, [[arXiv:1312.6320](#)].
- [116] T. Matsubara, *Recursive solutions of Lagrangian perturbation theory*, Phys. Rev. D **92** (July, 2015) 023534, [[arXiv:1505.01481](#)].
- [117] T. Buchert, *Lagrangian theory of gravitational instability of Friedman-Lemaitre cosmologies and the 'Zel'dovich approximation'*, MNRAS **254** (Feb., 1992) 729–737.
- [118] A. Barreira, T. Lazeyras, and F. Schmidt, *Galaxy bias from forward models: linear and second-order bias of IllustrisTNG galaxies*, J. Cosmology Astropart. Phys. **2021** (Aug., 2021) 029, [[arXiv:2105.02876](#)].
- [119] T. Lazeyras, A. Barreira, and F. Schmidt, *Assembly bias in quadratic bias parameters of dark matter halos from forward modeling*, J. Cosmology Astropart. Phys. **2021** (Oct., 2021) 063, [[arXiv:2106.14713](#)].
- [120] J. Mannseth, T. Selland Kleppe, and H. J. Skaug, *On the application of higher order symplectic integrators in Hamiltonian Monte Carlo*, arXiv e-prints (Aug., 2016).
- [121] M. Hernández-Sánchez, F.-S. Kitaura, M. Ata, and C. Dalla Vecchia, *Higher order Hamiltonian Monte Carlo sampling for cosmological large-scale structure analysis*, MNRAS **502** (Apr., 2021) 3976–3992, [[arXiv:1911.02667](#)].
- [122] Y. Feng, U. Seljak, and M. Zaldarriaga, *Exploring the posterior surface of the large scale structure reconstruction*, Journal of Cosmology and Astroparticle Physics **2018** (2018), no. 07 043.
- [123] A. Gelman and D. B. Rubin, *Inference from Iterative Simulation Using Multiple Sequences*, Statistical Science **7** (Jan., 1992) 457–472.
- [124] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, Bayesian data analysis. Chapman and Hall/CRC, 1995.

- [125] D. Vats and C. Knudson, *Revisiting the gelman–rubin diagnostic*, Statistical Science **36** (2021), no. 4 518–529, [[arXiv:1812.09384](#)].
- [126] K. J. Olejniczak, *The hartley transform*, The Transforms and Applications Handbook (2000) 281–330.
- [127] K. Jones, *The regularized fast Hartley transform: optimal formulation of real-data fast Fourier transform for silicon-based implementation in resource-constrained environments*. Springer Science & Business Media, 2010.
- [128] F. S. Kitaura and T. A. Enßlin, *Bayesian reconstruction of the cosmological large-scale structure: methodology, inverse algorithms and numerical optimization*, MNRAS **389** (Sept., 2008) 497–544, [[arXiv:0705.0429](#)].
- [129] A. Sokal, *Monte carlo methods in statistical mechanics: foundations and new algorithms*, in Functional integration, pp. 131–192. Springer, 1997.
- [130] N. Madras and A. D. Sokal, *The pivot algorithm: a highly efficient monte carlo method for the self-avoiding walk*, Journal of Statistical Physics **50** (1988), no. 1 109–186.
- [131] S. P. Brooks and A. Gelman, *General methods for monitoring convergence of iterative simulations*, Journal of computational and graphical statistics **7** (1998), no. 4 434–455.
- [132] D. J. Bartlett, A. Kostić, H. Desmond, J. Jasche, and G. Lavaux, *Constraints on dark matter annihilation and decay from the large-scale structure of the nearby universe*, Phys. Rev. D **106** (Nov., 2022) [[arXiv:2205.12916](#)].
- [133] A. Kostić, D. J. Bartlett, and H. Desmond, *No evidence for p - or d -wave dark matter annihilation from local large-scale structure*, arXiv e-prints (Apr., 2023) [[arXiv:2304.10301](#)].
- [134] L. Goodenough and D. Hooper, *Possible Evidence For Dark Matter Annihilation In The Inner Milky Way From The Fermi Gamma Ray Space Telescope*, arXiv e-prints (Oct., 2009) [[arXiv:0910.2998](#)].
- [135] M. Ajello et al., *Fermi-LAT Observations of High-Energy Gamma-Ray Emission toward the Galactic Center*, ApJ **819** (Mar., 2016) 44, [[arXiv:1511.02938](#)].
- [136] T. Linden, N. L. Rodd, B. R. Safdi, and T. R. Slatyer, *High-energy tail of the Galactic Center gamma-ray excess*, Phys. Rev. D **94** (Nov., 2016) [[arXiv:1604.01026](#)].

- [137] D. Hooper and L. Goodenough, *Dark matter annihilation in the Galactic Center as seen by the Fermi Gamma Ray Space Telescope*, Physics Letters B **697** (Mar., 2011) 412–428, [arXiv:1010.2752].
- [138] D. Hooper and T. Linden, *Origin of the gamma rays from the Galactic Center*, Phys. Rev. D **84** (Dec., 2011) [arXiv:1110.0006].
- [139] D. Hooper and T. R. Slatyer, *Two emission mechanisms in the Fermi Bubbles: A possible signal of annihilating dark matter*, Physics of the Dark Universe **2** (Sept., 2013) 118–138, [arXiv:1302.6589].
- [140] B. Zhou, Y.-F. Liang, X. Huang, et al., *GeV excess in the Milky Way: The role of diffuse galactic gamma-ray emission templates*, Phys. Rev. D **91** (June, 2015) [arXiv:1406.6948].
- [141] T. Daylan, D. P. Finkbeiner, D. Hooper, et al., *The characterization of the gamma-ray signal from the central Milky Way: A case for annihilating dark matter*, Physics of the Dark Universe **12** (June, 2016) 1–23, [arXiv:1402.6703].
- [142] I. Cholis, Y.-M. Zhong, S. D. McDermott, and J. P. Surdutovich, *Return of the templates: Revisiting the Galactic Center excess with multimessenger observations*, Phys. Rev. D **105** (May, 2022) [arXiv:2112.09706].
- [143] R. J. J. Grand and S. D. M. White, *Dark matter annihilation and the Galactic Centre Excess*, MNRAS **511** (Mar., 2022) L55–L59, [arXiv:2201.03567].
- [144] K. N. Abazajian, *The consistency of Fermi-LAT observations of the galactic center with a millisecond pulsar population in the central stellar cluster*, J. Cosmology Astropart. Phys. **2011** (Mar., 2011) 010, [arXiv:1011.4275].
- [145] R. M. O’Leary, M. D. Kistler, M. Kerr, and J. Dexter, *Young Pulsars and the Galactic Center GeV Gamma-ray Excess*, arXiv e-prints (Apr., 2015) [arXiv:1504.02477].
- [146] J. Petrović, P. D. Serpico, and G. Zaharijas, *Millisecond pulsars and the Galactic Center gamma-ray excess: the importance of luminosity function and secondary emission*, J. Cosmology Astropart. Phys. **2015** (Feb., 2015) 023, [arXiv:1411.2980].
- [147] S. K. Lee, M. Lisanti, B. R. Safdi, T. R. Slatyer, and W. Xue, *Evidence for Unresolved γ -Ray Point Sources in the Inner Galaxy*, Phys. Rev. Lett. **116** (Feb., 2016) 051103, [arXiv:1506.05124].

- [148] M. Buschmann, N. L. Rodd, B. R. Safdi, et al., *Foreground mismodeling and the point source explanation of the Fermi Galactic Center excess*, Phys. Rev. D **102** (July, 2020) 023023, [arXiv:2002.12373].
- [149] A. Gautam, R. M. Crocker, L. Ferrario, et al., *Millisecond pulsars from accretion-induced collapse as the origin of the Galactic Centre gamma-ray excess signal*, Nature Astronomy **6** (Apr., 2022) 703–707, [arXiv:2106.00222].
- [150] D. Hooper, I. Cholis, T. Linden, J. M. Siegal-Gaskins, and T. R. Slatyer, *Millisecond pulsars cannot account for the inner Galaxy’s GeV excess*, Phys. Rev. D **88** (Oct., 2013) 083009, [arXiv:1305.0830].
- [151] I. Cholis, D. Hooper, and T. Linden, *Challenges in explaining the Galactic Center gamma-ray excess with millisecond pulsars*, J. Cosmology Astropart. Phys. **2015** (June, 2015) 043, [arXiv:1407.5625].
- [152] R. K. Leane and T. R. Slatyer, *Revival of the Dark Matter Hypothesis for the Galactic Center Gamma-Ray Excess*, Phys. Rev. Lett. **123** (2019), no. 24 241101, [arXiv:1904.08430].
- [153] K. N. Abazajian and M. Kaplinghat, *Detection of a gamma-ray source in the Galactic Center consistent with extended emission from dark matter annihilation and concentrated astrophysical emission*, Phys. Rev. D **86** (Oct., 2012) 083511, [arXiv:1207.6047].
- [154] C. Gordon and O. Macías, *Dark matter and pulsar model constraints from Galactic Center Fermi-LAT gamma-ray observations*, Phys. Rev. D **88** (Oct., 2013) 083521, [arXiv:1306.5725].
- [155] K. N. Abazajian, N. Canac, S. Horiuchi, and M. Kaplinghat, *Astrophysical and dark matter interpretations of extended gamma-ray emission from the Galactic Center*, Phys. Rev. D **90** (July, 2014) 023526, [arXiv:1402.4090].
- [156] F. Calore, I. Cholis, and C. Weniger, *Background model systematics for the Fermi GeV excess*, J. Cosmology Astropart. Phys. **2015** (Mar., 2015) 038, [arXiv:1409.0042].
- [157] S. Horiuchi, M. Kaplinghat, and A. Kwa, *Investigating the uniformity of the excess gamma rays towards the galactic center region*, J. Cosmology Astropart. Phys. **2016** (Nov., 2016) 053, [arXiv:1604.01402].
- [158] J. Petrović, P. D. Serpico, and G. Zaharijaš, *Galactic Center gamma-ray “excess” from an active past of the Galactic Centre?*, J. Cosmology Astropart. Phys. **2014** (Oct., 2014) 052, [arXiv:1405.7928].

- [159] E. Carlson and S. Profumo, *Cosmic ray protons in the inner Galaxy and the Galactic Center gamma-ray excess*, Phys. Rev. D **90** (July, 2014) 023015, [arXiv:1405.7685].
- [160] I. Cholis, C. Evoli, F. Calore, et al., *The Galactic Center GeV excess from a series of leptonic cosmic-ray outbursts*, J. Cosmology Astropart. Phys. **2015** (Dec., 2015) 005, [arXiv:1506.05119].
- [161] D. Gaggero, M. Taoso, A. Urbano, M. Valli, and P. Ullio, *Towards a realistic astrophysical interpretation of the gamma-ray Galactic center excess*, J. Cosmology Astropart. Phys. **2015** (Dec., 2015) 056, [arXiv:1507.06129].
- [162] O. Macias, C. Gordon, R. M. Crocker, et al., *Galactic bulge preferred over dark matter for the Galactic centre gamma-ray excess*, Nature Astronomy **2** (Mar., 2018) 387–392, [arXiv:1611.06644].
- [163] R. Bartels, E. Storm, C. Weniger, and F. Calore, *The Fermi-LAT GeV excess as a tracer of stellar mass in the Galactic bulge*, Nature Astronomy **2** (Aug., 2018) 819–828, [arXiv:1711.04778].
- [164] Muon g-2 Collaboration, *Measurement of the Positive Muon Anomalous Magnetic Moment to 0.46 ppm*, Phys. Rev. Lett. **126** (Apr., 2021) 141801.
- [165] T. Aaltonen et al., *High-precision measurement of the W boson mass with the CDF II detector*, Science **376** (2022), no. 6589 170–176.
- [166] M. Kamionkowski, *WIMP and Axion Dark Matter*, in High Energy Physics and Cosmology, 1997 Summer School (E. Gava, A. Masiero, K. S. Narain, et al., eds.), vol. 14, p. 394, Jan., 1998. [arXiv:hep-ph/hep-ph/9710467].
- [167] G. Bertone, D. Hooper, and J. Silk, *Particle dark matter: evidence, candidates and constraints*, Phys. Rep. **405** (Jan., 2005) 279–390, [arXiv:hep-ph/hep-ph/0404175].
- [168] M. Ackermann et al., *Searching for Dark Matter Annihilation from Milky Way Dwarf Spheroidal Galaxies with Six Years of Fermi Large Area Telescope Data*, Phys. Rev. Lett. **115** (Dec., 2015) 231301, [arXiv:1503.02641].
- [169] A. Albert et al., *Searching for Dark Matter Annihilation in Recently Discovered Milky Way Satellites with Fermi-Lat*, ApJ **834** (Jan., 2017) 110, [arXiv:1611.03184].
- [170] V. Gammaldi, J. Pérez-Romero, J. Coronado-Blázquez, et al., *Dark matter search in dwarf irregular galaxies with the Fermi Large Area Telescope*, Phys. Rev. D **104** (Oct., 2021) 083026, [arXiv:2109.11291].

- [171] Y. G. Kim and K. Y. Lee, *Minimal model of fermionic dark matter*, Phys. Rev. D **75** (June, 2007) 115012, [[arXiv:hep-ph/0611069](#)].
- [172] K. Y. Lee, Y. G. Kim, and S. Shin, *Singlet fermionic dark matter*, Journal of High Energy Physics **2008** (May, 2008) 100, [[arXiv:0803.2932](#)].
- [173] J. Kumar and D. Marfatia, *Matrix element analyses of dark matter scattering and annihilation*, Phys. Rev. D **88** (July, 2013) 014035, [[arXiv:1305.1611](#)].
- [174] F. Giacchino, L. Lopez-Honorez, and M. H. G. Tytgat, *Scalar dark matter models with significant internal bremsstrahlung*, J. Cosmology Astropart. Phys. **2013** (Oct., 2013) 025, [[arXiv:1307.6480](#)].
- [175] T. Toma, *Internal Bremsstrahlung Signature of Real Scalar Dark Matter and Consistency with Thermal Relic Density*, Phys. Rev. Lett. **111** (Aug., 2013) 091301, [[arXiv:1307.6181](#)].
- [176] A. J. Cuesta, T. E. Jeltema, F. Zandanel, et al., *Dark Matter Decay and Annihilation in the Local Universe: Clues from Fermi*, ApJ **726** (Jan., 2011) L6, [[arXiv:1007.3469](#)].
- [177] S. Ando and E. Komatsu, *Anisotropy of the cosmic gamma-ray background from dark matter annihilation*, Phys. Rev. D **73** (2006) 023521, [[astro-ph/0512217](#)].
- [178] **Fermi-LAT** Collaboration, M. Ackermann et al., *Anisotropies in the diffuse gamma-ray background measured by the Fermi LAT*, Phys. Rev. D **85** (2012) 083007, [[arXiv:1202.2856](#)].
- [179] M. Fornasa et al., *Angular power spectrum of the diffuse gamma-ray emission as measured by the Fermi Large Area Telescope and constraints on its dark matter interpretation*, Phys. Rev. D **94** (2016), no. 12 123005, [[arXiv:1608.07289](#)].
- [180] M. Boylan-Kolchin, V. Springel, S. D. M. White, A. Jenkins, and G. Lemson, *Resolving cosmic structure formation with the Millennium-II Simulation*, MNRAS **398** (Sept., 2009) 1150–1164, [[arXiv:0903.3041](#)].
- [181] J. F. Navarro, A. Ludlow, V. Springel, et al., *The diversity and similarity of simulated cold dark matter haloes*, MNRAS **402** (Feb., 2010) 21–34, [[arXiv:0810.1522](#)].
- [182] V. Springel, J. Wang, M. Vogelsberger, et al., *The Aquarius Project: the subhaloes of galactic haloes*, MNRAS **391** (Dec., 2008) 1685–1711, [[arXiv:0809.0898](#)].
- [183] D. J. Bartlett, H. Desmond, and P. G. Ferreira, *Constraints on Galileons from the positions of supermassive black holes*, Phys. Rev. D **103** (Jan., 2021) 023523.

- [184] H. Desmond, M. L. Hutt, J. Devriendt, and A. Slyz, *Catalogues of voids as antihaloed in the local Universe*, MNRAS **511** (Mar., 2022) L45–L49, [arXiv:2109.09439].
- [185] M. L. Hutt, H. Desmond, J. Devriendt, and A. Slyz, *The effect of local Universe constraints on halo abundance and clustering*, MNRAS **516** (Nov., 2022) 3592–3601, [arXiv:2203.14724].
- [186] M. L. Hutt, H. Desmond, J. Devriendt, and A. Slyz, *The effect of local universe constraints on halo abundance and clustering*, <https://doi.org/10.5281/zenodo.5851241>, Mar., 2022.
- [187] G. Lavaux and J. Jasche, *Unmasking the masked Universe: the $2M++$ catalogue through Bayesian eyes*, MNRAS **455** (Jan., 2016) 3169–3179, [arXiv:1509.05040].
- [188] J. F. Navarro, C. S. Frenk, and S. D. M. White, *A Universal Density Profile from Hierarchical Clustering*, ApJ **490** (Dec., 1997) 493–508, [arXiv:astro-ph/astro-ph/9611107].
- [189] E. Piccirillo et al., *Velocity-dependent annihilation radiation from dark matter subhalos in cosmological simulations*, J. Cosmology Astropart. Phys. **2022** (Aug., 2022) 058, [arXiv:2203.08853].
- [190] G. Facchinetti, M. Stref, T. Lacroix, et al., *Analytical insight into dark matter subhalo boost factors for Sommerfeld-enhanced s - and p -wave γ -ray signals*, arXiv e-prints (Mar., 2022) [arXiv:2203.16491].
- [191] T. Lacroix et al., *Classification of gamma-ray targets for velocity-dependent and subhalo-boosted dark-matter annihilation*, J. Cosmology Astropart. Phys. **2022** (Oct., 2022) 021, [arXiv:2203.16440].
- [192] E. J. Baxter, J. Kumar, A. D. Paul, and J. Runburg, *Searching for velocity-dependent dark matter annihilation signals from extragalactic halos*, J. Cosmology Astropart. Phys. **2022** (Sept., 2022) 026, [arXiv:2205.02386].
- [193] M. Lisanti, S. Mishra-Sharma, N. L. Rodd, B. R. Safdi, and R. H. Wechsler, *Mapping extragalactic dark matter annihilation with galaxy surveys: A systematic study of stacked group searches*, Phys. Rev. D **97** (Mar., 2018) 063005.
- [194] R. Teyssier, *Cosmological hydrodynamics with adaptive mesh refinement. A new high resolution code called RAMSES*, A&A **385** (Apr., 2002) 337–364.

- [195] A. Bleuler, R. Teyssier, S. Carassou, and D. Martizzi, *PHEW: a parallel segmentation algorithm for three-dimensional AMR datasets. Application to structure detection in self-gravitating flows*, Computational Astrophysics and Cosmology **2** (June, 2015) 5, [[arXiv:1412.0510](#)].
- [196] J. J. Monaghan, *Smoothed particle hydrodynamics.*, ARA&A **30** (Jan., 1992) 543–574.
- [197] S. Colombi, M. J. Chodorowski, and R. Teyssier, *Cosmic velocity-gravity relation in redshift space*, MNRAS **375** (Feb., 2007) 348–370, [[arXiv:0805.1693](#)].
- [198] A. Okabe, B. Boots, K. Sugihara, and S. N. Chiu, *Spatial tessellations: concepts and applications of Voronoi diagrams*. John Wiley & Sons, 2000.
- [199] A. Charbonnier, C. Combet, and D. Maurin, *CLUMPY: A code for γ -ray signals from dark matter structures*, Computer Physics Communications **183** (Mar., 2012) 656–668, [[arXiv:1201.4728](#)].
- [200] V. Bonnivard, M. Hütten, E. Nezri, et al., *CLUMPY: Jeans analysis, γ -ray and ν fluxes from dark matter (sub-)structures*, Computer Physics Communications **200** (Mar., 2016) 336–349, [[arXiv:1506.07628](#)].
- [201] M. Hütten, C. Combet, and D. Maurin, *CLUMPY v3: γ -ray and ν signals from dark matter at all scales*, Computer Physics Communications **235** (Feb., 2019) 336–345, [[arXiv:1806.08639](#)].
- [202] L. Hernquist, *An Analytical Model for Spherical Galaxies and Bulges*, ApJ **356** (June, 1990) 359.
- [203] H. Zhao, *Analytical models for galactic nuclei*, MNRAS **278** (Jan., 1996) 488–496, [[arXiv:astro-ph/astro-ph/9509122](#)].
- [204] J. F. Navarro, E. Hayashi, C. Power, et al., *The inner structure of Λ CDM haloes - III. Universality and asymptotic slopes*, MNRAS **349** (Apr., 2004) 1039–1051, [[arXiv:astro-ph/astro-ph/0311231](#)].
- [205] B. Diemer, *Colossus: A python toolkit for cosmology, large-scale structure, and dark matter halos*, The Astrophysical Journal Supplement Series **239** (2018), no. 2 35.
- [206] L. Pieri, J. Lavalle, G. Bertone, and E. Branchini, *Implications of high-resolution simulations on indirect dark matter searches*, Phys. Rev. D **83** (Jan., 2011) 023518, [[arXiv:0908.0195](#)].

- [207] J. Lavalle, Q. Yuan, D. Maurin, and X. J. Bi, *Full calculation of clumpiness boost factors for antimatter cosmic rays in the light of Λ CDM N -body simulation results. Abandoning hope in clumpiness enhancement?*, *A&A* **479** (Feb., 2008) 427–452, [[arXiv:0709.3634](#)].
- [208] A. D. Ludlow, J. F. Navarro, V. Springel, et al., *The unorthodox orbits of substructure halos*, *The Astrophysical Journal* **692** (2009), no. 1 931.
- [209] M. Hütten, C. Combet, G. Maier, and D. Maurin, *Dark matter substructure modelling and sensitivity of the cherenkov telescope array to galactic dark halos*, *Journal of Cosmology and Astroparticle Physics* **2016** (2016), no. 09 047.
- [210] M. A. Sánchez-Conde and F. Prada, *The flattening of the concentration-mass relation towards low halo masses and its implications for the annihilation signal boost*, *MNRAS* **442** (Aug., 2014) 2271–2277, [[arXiv:1312.1729](#)].
- [211] J. S. Bullock, T. S. Kolatt, Y. Sigad, et al., *Profiles of dark haloes: evolution, scatter and environment*, *Monthly Notices of the Royal Astronomical Society* **321** (2001), no. 3 559–575.
- [212] M. A. Sánchez-Conde and F. Prada, *The flattening of the concentration–mass relation towards low halo masses and its implications for the annihilation signal boost*, *Monthly Notices of the Royal Astronomical Society* **442** (2014), no. 3 2271–2277.
- [213] R. H. Wechsler, J. S. Bullock, J. R. Primack, A. V. Kravtsov, and A. Dekel, *Concentrations of dark halos from their assembly histories*, *The Astrophysical Journal* **568** (2002), no. 1 52.
- [214] F. Prada, A. A. Klypin, A. J. Cuesta, J. E. Betancort-Rijo, and J. Primack, *Halo concentrations in the standard Λ cold dark matter cosmology*, *MNRAS* **423** (July, 2012) 3018–3030, [[arXiv:1104.5130](#)].
- [215] E. L. Lokas and G. A. Mamon, *Properties of spherical galaxies and clusters with an NFW density profile*, *MNRAS* **321** (Feb., 2001) 155–166, [[arXiv:astro-ph/0002395](#)].
- [216] K. K. Boddy, J. Kumar, and L. E. Strigari, *Effective J -factor of the Galactic Center for velocity-dependent dark matter annihilation*, *Phys. Rev. D* **98** (Sept., 2018) 063012, [[arXiv:1805.08379](#)].
- [217] L. Osipkov, *Spherical systems of gravitating bodies with ellipsoidal velocity distribution.*, *Pisma v Astronomicheskii Zhurnal* **5** (1979) 77–80.

- [218] D. Merritt, *Spherical stellar systems with spheroidal velocity distributions*, The astronomical journal **90** (1985) 1027–1037.
- [219] O. E. Gerhard, *A new family of distribution functions for spherical galaxies*, Monthly Notices of the Royal Astronomical Society **250** (1991), no. 4 812–830.
- [220] B. Boucher, J. Kumar, V. B. Le, and J. Runburg, *J-factors for velocity-dependent dark matter annihilation*, Physical Review D **106** (2022), no. 2 023025.
- [221] J. Binney and S. Tremaine, Galactic dynamics, vol. 13. Princeton university press, 2011.
- [222] A. Zonca, L. Singer, D. Lenz, et al. Journal of Open Source Software **4** (Mar., 2019) 1298.
- [223] K. M. Górski, E. Hivon, A. J. Banday, et al. ApJ **622** (Apr., 2005) 759–771.
- [224] M. Wood, R. Caputo, E. Charles, et al., *Fermipy: An open-source Python package for analysis of Fermi-LAT Data*, in 35th International Cosmic Ray Conference (ICRC2017), vol. 301 of International Cosmic Ray Conference, p. 824, Jan., 2017. [[arXiv:1707.09551](#)].
- [225] F. Acero et al., *Development of the Model of Galactic Interstellar Emission for Standard Point-source Analysis of Fermi Large Area Telescope Data*, ApJS **223** (Apr., 2016) 26, [[arXiv:1602.07246](#)].
- [226] A. Lewis, Oct., 2019.
- [227] T. E. Jeltema and S. Profumo, *Fitting the gamma-ray spectrum from dark matter with DMFIT: GLAST and the galactic center region*, J. Cosmology Astropart. Phys. **2008** (Nov., 2008) 003, [[arXiv:0808.2641](#)].
- [228] M. Cirelli, G. Corcella, A. Hektor, et al., *PPPC 4 DM ID: a poor particle physicist cookbook for dark matter indirect detection*, J. Cosmology Astropart. Phys. **2011** (Mar., 2011) 051, [[arXiv:1012.4515](#)].
- [229] D. Hashimoto, A. J. Nishizawa, and M. Takada, *Constraining Dark matter annihilation with Dark Energy Survey Y3 LSBG sample*, arXiv e-prints (Feb., 2022) [[arXiv:2202.01400](#)].
- [230] G. Schwarz, *Estimating the Dimension of a Model*, Annals of Statistics **6** (July, 1978) 461–464.
- [231] Y. Zhao, X.-J. Bi, H.-Y. Jia, P.-F. Yin, and F.-R. Zhu, *Constraint on the velocity dependent dark matter annihilation cross section from Fermi-LAT observations of dwarf galaxies*, Phys. Rev. D **93** (Apr., 2016) 083513, [[arXiv:1601.02181](#)].

- [232] K. K. Boddy, J. Kumar, A. B. Pace, J. Runburg, and L. E. Strigari, *Effective J -factors for Milky Way dwarf spheroidal galaxies with velocity-dependent annihilation*, Phys. Rev. D **102** (July, 2020) 023029, [arXiv:1909.13197].
- [233] M. Boudaud, T. Lacroix, M. Stref, and J. Lavalle, *Robust cosmic-ray constraints on p -wave annihilating MeV dark matter*, Phys. Rev. D **99** (Mar., 2019) 061302, [arXiv:1810.01680].
- [234] E. Branchini, S. Camera, A. Cuoco, et al., *Cross-correlating the γ -ray Sky with Catalogs of Galaxy Clusters*, ApJS **228** (Jan., 2017) 8, [arXiv:1612.05788].
- [235] M. Ackermann et al., *Unresolved Gamma-Ray Sky through its Angular Power Spectrum*, Phys. Rev. Lett. **121** (Dec., 2018) 241101, [arXiv:1812.02079].
- [236] M. Fornasa, J. Zavala, M. A. Sánchez-Conde, et al., *Characterization of dark-matter-induced anisotropies in the diffuse gamma-ray background*, Monthly Notices of the Royal Astronomical Society **429** (2013), no. 2 1529–1553.
- [237] J. Lavalle, J. Pochon, P. Salati, and R. Taillet, *Clumpiness of dark matter and the positron annihilation signal*, Astronomy & Astrophysics **462** (2007), no. 3 827–840.
- [238] G. R. Blumenthal, S. M. Faber, R. Flores, and J. R. Primack, *Contraction of Dark Matter Galactic Halos Due to Baryonic Infall*, ApJ **301** (Feb., 1986) 27.
- [239] O. Y. Gnedin, A. V. Kravtsov, A. A. Klypin, and D. Nagai, *Response of Dark Matter Halos to Condensation of Baryons: Cosmological Simulations and Improved Adiabatic Contraction Model*, ApJ **616** (Nov., 2004) 16–26, [arXiv:astro-ph/astro-ph/0406247].
- [240] A. Pontzen and F. Governato, *How supernova feedback turns dark matter cusps into cores*, Mon. Not. Roy. Astron. Soc. **421** (Apr., 2012) 3464–3471, [arXiv:1106.0499].
- [241] A. Del Popolo and F. Pace, *The Cusp/Core problem: supernovae feedback versus the baryonic clumps and dynamical friction model*, Astrophys. Space Sci. **361** (2016), no. 5 162, [arXiv:1502.01947]. [Erratum: Astrophys.Space Sci. 361, 225 (2016)].
- [242] E. Board et al., *Velocity-dependent J -factors for annihilation radiation from cosmological simulations*, J. Cosmology Astropart. Phys. **2021** (Apr., 2021) 070, [arXiv:2101.06284].

- [243] D. McKeown et al., *Amplified J -factors in the Galactic Centre for velocity-dependent dark matter annihilation in FIRE simulations*, MNRAS **513** (June, 2022) 55–70, [arXiv:2111.03076].
- [244] A. Franceschini, G. Rodighiero, and M. Vaccari, *The extragalactic optical-infrared background radiations, their time evolution and the cosmic photon-photon opacity* <https://doi.org/10.1051/0004-6361/200809691> *astron.*, Astrophys **487** (2008), no. 837 0805–1841.
- [245] A. Franceschini and G. Rodighiero, *The extragalactic background light revisited and the cosmic photon-photon opacity*, Astronomy & Astrophysics **603** (2017) A34.
- [246] T. R. Slatyer, *Les Houches Lectures on Indirect Detection of Dark Matter*, arXiv e-prints (Sept., 2021) [arXiv:2109.02696].
- [247] P. D. Serpico, E. Sefusatti, M. Gustafsson, and G. Zaharijas, *Extragalactic gamma-ray signal from dark matter annihilation: a power spectrum based computation*, Monthly Notices of the Royal Astronomical Society: Letters **421** (2012), no. 1 L87–L91.
- [248] M. Hütten, C. Combet, and D. Maurin, *Extragalactic diffuse γ -rays from dark matter annihilation: revised prediction and full modelling uncertainties*, J. Cosmology Astropart. Phys. **2018** (Feb., 2018) 005, [arXiv:1711.08323].
- [249] W. H. Press and P. Schechter, *Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation*, ApJ **187** (Feb., 1974) 425–438.
- [250] R. Piffaretti, M. Arnaud, G. Pratt, E. Pointecouteau, and J.-B. Melin, *The *mcxc*: a meta-catalogue of x-ray detected clusters of galaxies*, Astronomy & Astrophysics **534** (2011) A109.
- [251] D. Merritt, *The Distribution of Dark Matter in the Coma Cluster*, ApJ **313** (Feb., 1987) 121.
- [252] A. E. Evrard, C. A. Metzler, and J. F. Navarro, *Mass Estimates of X-Ray Clusters*, ApJ **469** (Oct., 1996) 494, [arXiv:astro-ph/9510058].
- [253] R. A. Sunyaev and Y. B. Zeldovich, *The Spectrum of Primordial Radiation, its Distortions and their Significance*, Comments on Astrophysics and Space Physics **2** (Mar., 1970) 66.
- [254] R. A. Sunyaev and I. B. Zeldovich, *Microwave background radiation as a probe of the contemporary structure and history of the universe*, ARA&A **18** (Jan., 1980) 537–560.

- [255] H. Bonnet, Y. Mellier, and B. Fort, *First Detection of a Gravitational Weak Shear at the Periphery of CL 0024+1654*, ApJ **427** (June, 1994) L83.
- [256] G. Fahlman, N. Kaiser, G. Squires, and D. Woods, *Dark Matter in MS 1224 from Distortion of Background Galaxies*, ApJ **437** (Dec., 1994) 56, [[arXiv:astro-ph/9402017](#)].
- [257] S. Stopyra, H. V. Peiris, A. Pontzen, J. Jasche, and P. Natarajan, *Quantifying the rarity of the local super-volume*, MNRAS **507** (Nov., 2021) 5425–5431, [[arXiv:2107.06903](#)].
- [258] S. Tassev, M. Zaldarriaga, and D. J. Eisenstein, *Solving large scale structure in ten easy steps with COLA*, J. Cosmology Astropart. Phys. **2013** (June, 2013) 036, [[arXiv:1301.0322](#)].
- [259] D. Blas, J. Lesgourgues, and T. Tram, *The Cosmic Linear Anisotropy Solving System (CLASS). Part II: Approximation schemes*, Journal of Cosmology and Astro-Particle Physics **2011** (Jul, 2011) 034, [[arXiv:1104.2933](#)].
- [260] D. J. Eisenstein and W. Hu, *Baryonic Features in the Matter Transfer Function*, ApJ **496** (Mar., 1998) 605–614, [[arXiv:astro-ph/9709112](#)].
- [261] S. Stopyra, H. V. Peiris, A. Pontzen, J. Jasche, and G. Lavaux, *Towards accurate field-level inference of massive cosmic structures*, 2023.
- [262] E. Hairer and G. Wanner, Solving Ordinary Differential Equations II. Stiff and Differential-Algebraic Problems, vol. 14. Springer Berlin Heidelberg New York, 01, 1996.
- [263] E. W. Kolb and M. S. Turner, The Early Universe, vol. 294. Nature Publishing Group UK London, 1981.
- [264] S. Campbell, B. Dutta, and E. Komatsu, *Effects of velocity-dependent dark matter annihilation on the energy spectrum of the extragalactic gamma-ray background*, Phys. Rev. D **82** (Nov., 2010) 095007, [[arXiv:1009.3530](#)].
- [265] S. Campbell and B. Dutta, *Effects of p-wave annihilation on the angular power spectrum of extragalactic gamma-rays from dark matter annihilation*, Phys. Rev. D **84** (Oct., 2011) 075004, [[arXiv:1106.4621](#)].
- [266] K. Griest and D. Seckel, *Three exceptions in the calculation of relic abundances*, Phys. Rev. D **43** (May, 1991) 3191–3203.
- [267] K. Jedamzik and M. Pospelov, *Big Bang nucleosynthesis and particle dark matter*, New Journal of Physics **11** (Oct., 2009) 105028, [[arXiv:0906.2087](#)].

- [268] V. Poulin and P. D. Serpico, *Nonuniversal BBN bounds on electromagnetically decaying particles*, Phys. Rev. D **91** (May, 2015) 103007, [arXiv:1503.04852].
- [269] S. Galli, F. Iocco, G. Bertone, and A. Melchiorri, *CMB constraints on dark matter models with large annihilation cross section*, Phys. Rev. D **80** (July, 2009) 023505, [arXiv:0905.0003].
- [270] M. Kawasaki, H. Nakatsuka, K. Nakayama, and T. Sekiguchi, *Revisiting CMB constraints on dark matter annihilation*, J. Cosmology Astropart. Phys. **2021** (Dec., 2021) 015, [arXiv:2105.08334].
- [271] T. R. Slatyer, *Indirect dark matter signatures in the cosmic dark ages. I. Generalizing the bound on s -wave dark matter annihilation from Planck results*, Phys. Rev. D **93** (Jan., 2016) 023527, [arXiv:1506.03811].
- [272] R. Diamanti, L. Lopez-Honorez, O. Mena, S. Palomares-Ruiz, and A. C. Vincent, *Constraining dark matter late-time energy injection: decays and p -wave annihilations*, J. Cosmology Astropart. Phys. **2014** (Feb., 2014) 017, [arXiv:1308.2578].
- [273] H. Liu, W. Qin, G. W. Ridgway, and T. R. Slatyer, *Lyman- α constraints on cosmic heating from dark matter annihilation and decay*, Phys. Rev. D **104** (Aug., 2021) 043514, [arXiv:2008.01084].
- [274] J. Choquette, J. M. Cline, and J. M. Cornell, *p -wave annihilating dark matter from a decaying predecessor and the Galactic Center excess*, Phys. Rev. D **94** (July, 2016) 015018, [arXiv:1604.01039].
- [275] C. Johnson, R. Caputo, C. Karwin, et al., *Search for gamma-ray emission from p -wave dark matter annihilation in the Galactic Center*, Phys. Rev. D **99** (May, 2019) 103007, [arXiv:1904.06261].
- [276] K. Kiriū, J. Kumar, and J. Runburg, *The Velocity-Dependent J -factor of the Milky Way Halo: Does What Happens in the Galactic Bulge Stay in the Galactic Bulge?*, arXiv e-prints (Aug., 2022) [arXiv:2208.14002].
- [277] M. Ackermann et al., *Constraints on the Galactic Halo Dark Matter from Fermi-LAT Diffuse Measurements*, ApJ **761** (Dec., 2012) 91, [arXiv:1205.6474].
- [278] L. J. Chang, M. Lisanti, and S. Mishra-Sharma, *Search for dark matter annihilation in the Milky Way halo*, Phys. Rev. D **98** (Dec., 2018) 123004, [arXiv:1804.04132].

- [279] M. Lisanti, S. Mishra-Sharma, N. L. Rodd, and B. R. Safdi, *Search for Dark Matter Annihilation in Galaxy Groups*, Phys. Rev. Lett. **120** (Mar., 2018) 101101, [[arXiv:1708.09385](#)].
- [280] M. A. Sánchez-Conde, M. Cannoni, F. Zandanel, M. E. Gómez, and F. Prada, *Dark matter searches with Cherenkov telescopes: nearby dwarf galaxies or local galaxy clusters?*, J. Cosmology Astropart. Phys. **2011** (Dec., 2011) 011, [[arXiv:1104.3530](#)].
- [281] S. Ando, A. Geringer-Sameth, N. Hiroshima, et al., *Structure formation models weaken limits on WIMP dark matter from dwarf spheroidal galaxies*, Phys. Rev. D **102** (Sept., 2020) 061302, [[arXiv:2002.11956](#)].
- [282] M. Regis, J. Reynoso-Cordova, M. D. Filipović, et al., *The EMU view of the Large Magellanic Cloud: troubles for sub-TeV WIMPs*, J. Cosmology Astropart. Phys. **2021** (Nov., 2021) 046, [[arXiv:2106.08025](#)].
- [283] D. Hashimoto, A. J. Nishizawa, M. Takada, and O. Macias, *Dark matter constraint with gamma-ray galaxies cross correlation scales with N* , arXiv e-prints (Sept., 2021) [[arXiv:2109.08832](#)].
- [284] S. Ammazzalorso et al., *Detection of Cross-Correlation between Gravitational Lensing and γ Rays*, Phys. Rev. Lett. **124** (Mar., 2020) 101102, [[arXiv:1907.13484](#)].
- [285] Y. Zhao, X.-J. Bi, P.-F. Yin, and X. Zhang, *Constraint on the velocity dependent dark matter annihilation cross section from gamma-ray and kinematic observations of ultrafaint dwarf galaxies*, Phys. Rev. D **97** (Mar., 2018) 063013, [[arXiv:1711.04696](#)].
- [286] E. J. Baxter, J. Kumar, A. B. Pace, and J. Runburg, *Prospects for measuring dark matter microphysics with observations of dwarf spheroidal galaxies*, J. Cosmology Astropart. Phys. **2021** (July, 2021) 030, [[arXiv:2103.11646](#)].
- [287] T. Cohen, K. Murase, N. L. Rodd, B. R. Safdi, and Y. Soreq, *γ -ray Constraints on Decaying Dark Matter and Implications for IceCube*, Phys. Rev. Lett. **119** (July, 2017) 021102, [[arXiv:1612.05638](#)].
- [288] L. A. Anchordoqui et al., *Hunting super-heavy dark matter with ultra-high energy photons*, Astroparticle Physics **132** (Nov., 2021) 102614, [[arXiv:2105.12895](#)].
- [289] D. Wadekar and Z. Wang, *Strong constraints on decay and annihilation of dark matter from heating of gas-rich dwarf galaxies*, Phys. Rev. D **106** (Oct., 2022) 075007, [[arXiv:2111.08025](#)].

- [290] S. Mau et al., *Milky Way Satellite Census. IV. Constraints on Decaying Dark Matter from Observations of Milky Way Satellite Galaxies*, ApJ **932** (June, 2022) 128, [arXiv:2201.11740].
- [291] H. Dejonghe, *Stellar dynamics and the description of stellar systems.*, Phys. Rep. **133** (Jan., 1986) 217–313.
- [292] A. Kostić, J. Jasche, D. K. Ramanah, and G. Lavaux, *Optimal machine-driven acquisition of future cosmological data*, A&A **657** (Jan., 2022) L17, [arXiv:2107.00657].
- [293] M. Tegmark et al., *The Three-Dimensional Power Spectrum of Galaxies from the Sloan Digital Sky Survey*, ApJ **606** (May, 2004) 702–740, [astro-ph/0310725].
- [294] W. J. Percival, R. C. Nichol, D. J. Eisenstein, et al., *The Shape of the Sloan Digital Sky Survey Data Release 5 Galaxy Power Spectrum*, ApJ **657** (Mar., 2007) 645–663, [arXiv:astro-ph/astro-ph/0608636].
- [295] A. Porredon, M. Crocce, J. Elvin-Poole, et al., *Dark Energy Survey Year 3 results: Cosmological constraints from galaxy clustering and galaxy-galaxy lensing using the MAGLIM lens sample*, Phys. Rev. D **106** (Nov., 2022) 103530, [arXiv:2105.13546].
- [296] T. Buchert, A. L. Melott, and A. G. Weiss, *Testing higher-order Lagrangian perturbation theory against numerical simulations I. Pancake models*, A&A **288** (Aug., 1994) 349–364.
- [297] F. R. Bouchet, S. Colombi, E. Hivon, and R. Juszkiewicz, *Perturbative Lagrangian approach to gravitational instability.*, A&A **296** (Apr., 1995) 575.
- [298] L.-Y. Chiang and P. Coles, *Phase information and the evolution of cosmological density perturbations*, MNRAS **311** (Feb., 2000) 809–824, [arXiv:astro-ph/astro-ph/9905250].
- [299] J. Byun, F. O. Franco, C. Howlett, C. Bonvin, and D. Obreschkow, *Constraining the growth rate of structure with phase correlations*, MNRAS **497** (Sept., 2020) 1765–1790, [arXiv:2005.06325].
- [300] N. Porqueres, O. Hahn, J. Jasche, and G. Lavaux, *A hierarchical field-level inference approach to reconstruction from sparse Lyman- α forest data*, A&A **642** (Oct., 2020) A139, [arXiv:2005.12928].
- [301] N. Porqueres, A. Heavens, D. Mortlock, and G. Lavaux, *Bayesian forward modelling of cosmic shear data*, MNRAS **502** (Apr., 2021) 3035–3044, [arXiv:2011.07722].

- [302] J. Einasto, G. Hütsi, E. Saar, et al., *Wavelet analysis of the cosmic web formation*, A&A **531** (July, 2011) A75, [[arXiv:1012.3550](#)].
- [303] Ž. Ivezić, S. M. Kahn, J. A. Tyson, et al., *LSST: From Science Drivers to Reference Design and Anticipated Data Products*, ApJ **873** (Mar., 2019) 111, [[arXiv:0805.2366](#)].
- [304] G. D. Racca, R. Laureijs, L. Stagnaro, et al., *The Euclid mission design*, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, vol. 9904 of Proc. SPIE, p. 99040O, July, 2016. [[arXiv:1610.05508](#)].
- [305] G. Lavaux and M. J. Hudson, *The $2M++$ galaxy redshift catalogue*, MNRAS **416** (Oct., 2011) 2840–2856, [[arXiv:1105.6107](#)].
- [306] R. A. Fisher, Theory of statistical estimation. Cambridge University Press, 1925.
- [307] F. Villaescusa-Navarro, D. Anglés-Alcázar, S. Genel, et al., *The CAMELS project: Cosmology and Astrophysics with MachinE Learning Simulations*, ArXiv e-prints (Oct., 2020) [[arXiv:2010.00619](#)].
- [308] F. Leclercq, G. Lavaux, J. Jasche, and B. Wandelt, *Comparing cosmic web classifiers using information theory*, J. Cosmology Astropart. Phys. **2016** (Aug., 2016) 027, [[arXiv:1606.06758](#)].
- [309] T. Y. Lam, T. Nishimichi, F. Schmidt, and M. Takada, *Testing Gravity with the Stacked Phase Space around Galaxy Clusters*, Phys. Rev. Lett. **109** (Aug., 2012) 051301, [[arXiv:1202.4501](#)].
- [310] N. Porqueres, J. Jasche, G. Lavaux, and T. Enßlin, *Inferring high-redshift large-scale structure dynamics from the Lyman- α forest*, A&A **630** (Oct, 2019) A151, [[arXiv:1907.02973](#)].
- [311] K. Koyama, *Cosmological tests of modified gravity*, Reports on Progress in Physics **79** (Apr., 2016) 046902, [[arXiv:1504.04623](#)].
- [312] G.-B. Zhao, M. Raveri, L. Pogosian, et al., *Dynamical dark energy in light of the latest observations*, Nature Astronomy **1** (Aug., 2017) 627–632, [[arXiv:1701.08165](#)].
- [313] E. D. Carlson, M. E. Machacek, and L. J. Hall, *Self-interacting Dark Matter*, ApJ **398** (Oct., 1992) 43.
- [314] W. Hu, R. Barkana, and A. Gruzinov, *Fuzzy Cold Dark Matter: The Wave Properties of Ultralight Particles*, Phys. Rev. Lett. **85** (Aug., 2000) 1158–1161, [[arXiv:astro-ph/astro-ph/0003365](#)].

- [315] F. Villaescusa-Navarro, F. Marulli, M. Viel, et al., *Cosmology with massive neutrinos I: towards a realistic modeling of the relation between matter, haloes and galaxies*, J. Cosmology Astropart. Phys. **2014** (Mar., 2014) 011, [[arXiv:1311.0866](#)].
- [316] J. P. Huchra, L. M. Macri, K. L. Masters, et al., *The 2MASS Redshift Survey—Description and Data Release*, ApJS **199** (Apr., 2012) 26, [[arXiv:1108.0669](#)].
- [317] D. H. Jones, M. A. Read, W. Saunders, et al., *The 6dF Galaxy Survey: final redshift release (DR3) and southern large-scale structures*, MNRAS **399** (Oct., 2009) 683–698, [[arXiv:0903.5451](#)].
- [318] K. N. Abazajian, J. K. Adelman-McCarthy, M. A. Agüeros, et al., *The Seventh Data Release of the Sloan Digital Sky Survey*, ApJS **182** (June, 2009) 543–558, [[arXiv:0812.0649](#)].
- [319] W. Saunders, W. J. Sutherland, S. J. Maddox, et al., *The PSCz catalogue*, MNRAS **317** (Sept., 2000) 55–63, [[astro-ph/0001117](#)].
- [320] M. F. Skrutskie, R. M. Cutri, R. Stiening, et al., *The Two Micron All Sky Survey (2MASS)*, AJ **131** (Feb., 2006) 1163–1183.
- [321] L. D. Landau and E. M. Lifshitz, *Theoretical physics, vol. 6, fluid mechanics*, 1987.
- [322] L. Landau and E. Lifshitz, *Theoretical physics, vol 2., the classical theory of fields*, Pergamon Press, Oxford (1971).
- [323] Y. B. Zel’dovich, *Gravitational instability: An approximate theory for large density perturbations.*, A&A **5** (Mar., 1970) 84–89.
- [324] S. Tassev and M. Zaldarriaga, *Estimating CDM particle trajectories in the mildly non-linear regime of structure formation. Implications for the density field in real and redshift space*, J. Cosmology Astropart. Phys. **12** (Dec., 2012) 11, [[arXiv:1203.5785](#)].
- [325] E. Bertschinger, *Simulations of Structure Formation in the Universe*, ARA&A **36** (Jan., 1998) 599–654.
- [326] T. Matsubara, *Nonlinear perturbation theory with halo bias and redshift-space distortions via the Lagrangian picture*, Phys. Rev. D **78** (Oct., 2008) 083519, [[arXiv:0807.1733](#)].

- [327] T. Buchert and J. Ehlers, *Averaging inhomogeneous Newtonian cosmologies.*, *A&A* **320** (Apr., 1997) 1–7, [[arXiv:astro-ph/astro-ph/9510056](#)].
- [328] S. Tassev and M. Zaldarriaga, *The mildly non-linear regime of structure formation*, *J. Cosmology Astropart. Phys.* **2012** (Apr., 2012) 013, [[arXiv:1109.4939](#)].
- [329] B. Stroustrup, *The C++ programming language fourth edition*, 2013.
- [330] D. Phan, N. Pradhan, and M. Jankowiak, *Composable effects for flexible and accelerated probabilistic programming in numpyro*, [arXiv:1912.11554](#).
- [331] E. Bingham et al., *Pyro: Deep universal probabilistic programming*, *J. Mach. Learn. Res.* **20** (2019) 28:1–28:6.
- [332] C. R. Harris et al., *Array programming with NumPy*, *Nature* **585** (Sept., 2020) 357–362.
- [333] P. Virtanen et al., *SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python*, *Nature Methods* **17** (2020) 261–272.
- [334] D. Foreman-Mackey *The Journal of Open Source Software* **1** (June, 2016) 24.

ACKNOWLEDGMENTS

This thesis was a result of cumulative work produced over the past two and a half years. Over this period, a lot of contribution and insight was generated through various discussions. I would foremost like to thank my supervisor Fabian Schmidt, who above all offered his patience and understanding for my spread interest and bifurcated focus over various topics. His numerous papers on bias modelling and effective field theory of large-scale structure, as well as his book on cosmology [42] were an enormous source of useful information from which I have learned a lot. I would like to thank him for the trust he put in my capabilities, offering me a PhD position within a field I have barely known existed beforehand. His persistence and vast knowledge are always an inspiration. I also thank Eiichiro Komatsu for supporting my curiosity and interest in various topics I have hung onto during the course of my PhD.

Next, I would like to thank all members of the Aquila consortium ¹, for not only providing me a lot of new project ideas and problems to work on, but amazing people to work with as well. Without the consortium, the results of Chapter 5 and Chapter 4 would not have been possible. For this I would in particular like to thank Jens Jasche, Guilhem Lavaux and Doogesh Kodi-Ramanah for providing me with guidance and useful discussions for my first paper presented in Chapter 5. Furthermore, I would like to thank Deaglan Bartlett and Harry Desmond for actively collaborating on the results of Chapter 4. For this, I also give my gratitude to the original developers of the CLUMPY package – Moritz Hütten, Céline Combet and David Maurin.

Also, I thank all the members of the physical cosmology group as well as my colleagues from the *International Max Planck Research School* for making this a more fun journey. I would especially like to thank Julia Stadler, Beatriz Tucci and Ivana Babić for offering plentiful discussions and their persistence in learning all about the EFTofLSS approach during our Friday book club meetings. I also thank Nguyen Nhat Minh for

¹<https://www.aquila-consortium.org>

useful discussions during the work on the results of Chapter 3 as well as his investment to the galactic dynamics book club from which I learned a lot.

I would also like to thank Andrej Obuljen, Marko Simonović, Uroš Seljak, Matias Zaldariagga, Henrique Rubira for useful discussions during the Šmartno 2022 conference, as well as Philipp Frank for discussion on the results of Appendix 3.8.2 and Florent Leclercq for the discussion on higher-order symplectic integrator schemes for HMC.

I also thank Jonathan Patterson for smoothly running the Glamdring Cluster at the University of Oxford, where most of the computation needed for Chapter 4 was performed. Alongside this, I thank also Martin Reinecke who helped in optimizing and developing the `LEFTfield` code and for his immense coding experience which always provided him with an answer to any question I would pose him. I am also grateful to him for suggesting me useful reading material on modern C++ programming [329], which I found surprisingly useful even with the vast resources available online.

I acknowledge support from the Starting Grant (ERC-2015-STG 678652) “GrInflaGal” of the European Research Council at MPA as well as the Dositeja Fund of the Serbian Ministry of Science.

I would also like to thank to the developers of the software that was heavily used for the papers presented in this thesis `healpy` and `HEALPix` [222, 223], `clumpy` [199–201], `numpyro` [330, 331], `emcee` [87], `numpy` [332], `scipy` [333], `fermipy` [224] and `corner` [334] packages.

Finally, I offer unfeigned gratitude to my life-long partner, Jovana Todorović, whose love and understanding provided me with interminable foundation. I also thank my family, for supporting every step of my journey into science, even when this lead me ever further from home.