

Department of Physics and Astronomy

Heidelberg University

Master thesis

in Physics

submitted by

Benedikt Schosser

born in München, Germany

2023

21cm Parameter Inference

Constraining extended Starobinsky Inflation and more with
future cosmological Experiments

This Master thesis has been carried out by Benedikt Schosser

at the

Institute for Theoretical Physics

under the supervision of

Prof. Dr. Tilman Plehn

21cm-Parameterinferenz – Einschränkung der erweiterten Starobinsky-Inflation und mehr mit zukünftigen kosmologischen Experimenten:

Starobinski-Inflation ist ein grundlegendes Modell, das auf einer Modifikation der allgemeinen Relativitätstheorie beruht und konsistent mit den Messungen des kosmischen Mikrowellenhintergrunds (CMB) der Planck-Mission ist. Eine zusätzliche Erweiterung auf Krümmungsterme höherer Ordnung ermöglicht die Untersuchung von Quantengravitationseffekten. Die Auswirkungen der Inflation lassen sich in den Temperaturanisotropien des CMB und der großräumigen Struktur beobachten. Das zukünftige Square Kilometer Array (SKA) wird die Verteilung von neutralem Wasserstoff im Universum über einen großen Bereich von Rotverschiebungen kartieren. Diese Karten können verwendet werden, um astrophysikalische Parameter mit Hilfe von Deep-Learning-Techniken für Likelihood freie Schlussfolgerungen zu bestimmen. Darüber hinaus zeigt diese Arbeit, dass SKA in Kombination mit zukünftigen CMB-Daten die derzeitigen Beschränkungen der Parameter des erweiterten Starobinski-Modells erheblich verbessern kann.

21cm Parameter Inference – Constraining extended Starobinsky Inflation and more with future cosmological Experiments:

Starobinsky inflation is a fundamental model, based on a modification of general relativity, consistent with the measurement of the Cosmic Microwave Background (CMB) by Planck. A further extension to higher order curvature terms allows to probe quantum gravity effects. The effects of inflation can be observed in the temperature anisotropies of the CMB and large scale structure. The future Square Kilometer Array (SKA) will map out the distribution of neutral hydrogen in the universe over a large range of redshifts. These maps can be used to determine astrophysical parameters using deep learning techniques for likelihood-free inference. In addition, this thesis shows that SKA, in combination with future CMB data, can significantly improve the current constraints on the parameters of the extended Starobinsky model.

Contents

1	Introduction	1
2	FLRW Universe	3
2.1	Geometry	3
2.2	Kinematics	4
2.3	Redshift	4
2.4	Dynamics	5
3	Inflation	7
3.1	Horizon Problem	7
3.2	Solution to the Horizon Problem	8
3.3	Equivalent Ways of Characterising Inflation	10
3.4	The Physics of Inflation	10
3.5	Reheating	12
4	Cosmological Perturbation Theory	13
4.1	Generalities	13
4.1.1	Perturbed Spacetime	13
4.1.2	Perturbed Matter	14
4.1.3	Perturbed Equations	15
4.1.4	Conserved Curvature Perturbation	16
4.2	Initial Conditions from Inflation	17
4.2.1	From Quantum to Classical	17
4.2.2	Mukhanov-Sasaki Equation	18
4.2.3	Quantum Fluctuations in de Sitter Space	19
4.2.4	Primordial Perturbations from Inflation	20
4.2.5	Gravitational Waves	21
5	Starobinsky Model of Inflation and its Extension	23
5.1	Formalism	23
5.2	Observables in Starobinsky Inflation	25
5.3	Field Dynamics in Starobinsky Inflation	27
5.4	Primordial Power Spectrum	29
5.5	R^3 -Extension	30

6	Contact with Observations	33
6.1	Cosmic Microwave Background	33
6.1.1	Statistics of Temperature Anisotropies	33
6.1.2	Transfer Function and Projection Effects	34
6.1.3	CMB Polarisation	35
6.1.4	Coherent Phases and Superhorizon Fluctuations	36
6.2	21cm Intensity Mapping	36
7	Statistics and Tools	41
7.1	Bayesian Inference	41
7.2	Metropolis-Hastings algorithm	42
7.3	Toy Example	43
7.4	MontePython	44
7.5	Likelihoods	46
7.5.1	CMB	46
7.5.2	Square Kilometer Array	47
7.6	CLASS	48
8	Results and Forecasts	50
8.1	Planck Results	51
8.2	CMB Forecasts	51
8.3	Square Kilometer Array Forecasts	55
9	21cm Inference with Deep Learning	59
9.1	Dataset	59
9.2	Basics of Deep Learning	60
9.2.1	Neural Network	60
9.2.2	Convolutional Neural Network	61
9.3	BayesFlow	62
9.3.1	Finding the Posterior	63
9.3.2	Composing Invertible Networks	64
9.4	Architecture and Training	65
9.4.1	Architecture	65
9.4.2	Training	66
9.4.3	Performance Validation	68
9.5	Results	69
10	Summary and Conclusion	76
A	Hubble Slow-Roll Projections	79
	Bibliography	83

CHAPTER 1

Introduction

Inflation was first proposed to give an elegant solution to some of the problems, e.g. the horizon problem, in the standard model of cosmology [1, 2, 3]. It is a very short period of accelerated expansion after the beginning of the Universe. This rapid expansion can explain the isotropy in the Cosmic Microwave Background (CMB), even in regions which would have never been in causal contact, in the standard Big Bang scenario [4]. Additionally, naturally occurring quantum fluctuations during inflation are the seeds for the structure observed in the Universe today. They are visible in the CMB [5] as well as in large-scale structures like galaxy clustering or the distribution of dark matter. Since inflation happened almost immediately after the Big Bang its condition can hardly be reproduced in experiments. Hence, the effects can only be tested by looking at its influence on structure formation in the sky and comparing it to theoretical predictions.

This thesis is focused on one particularly successful model, the Starobinsky or R^2 -inflationary model [3, 6, 7, 8]. According to the analysis by the Planck collaboration, it is one of the best-fit model to the data coming from the early Universe [5], the CMB. This model extends General Relativity (GR) by not only having the Ricci-scalar R but additionally a term proportional to R^2 in the action. It has one free parameter M , which manifests itself in the amplitude of the primordial scalar power spectrum and can be adjusted to fit the reported value. Naturally, this model predicts the measured scalar index n_s . From a phenomenological point of view it is highly motivated to also start including higher order terms. R^2 is merely a correction to GR, hence there is no ad hoc reason why a suppressed R^3 term should not be included. Now, the theory has a second parameter c parametrising the strength of the R^3 correction term and the goal of this thesis is to forecast constraints on these two Starobinsky parameters.

In this work future CMB experiments are used to test this extended Starobinsky model. On the one hand is the *LiteBIRD* satellite [9, 10, 11], which will detect B-mode polarisation with moderate resolution, but excellent sensitivity. Inflation predicts non vanishing tensor modes, leading to B-modes in the CMB. On the other hand, *CMB-S4* [12, 13, 14, 15], a ground-based detector with excellent resolution and sensitivity, is considered. Both of these experiments will start operating at the end of this decade. In addition to the data from the early Universe, there is a lot of

information about the inflationary period in the in the distribution of dark matter and hydrogen following it. Therefore, the future 21cm experiment *Square Kilometer Array* (SKA), which will map out the hydrogen in the Universe, can tighten the constraints significantly [16, 17, 18]. It is going to measure the 21cm-hyperfine transition line, giving insight into the epoch of reionization and the dark ages.

In Ch. 2 the basics of cosmology are explained before giving an introduction to general single-field inflation and its motivation in Ch. 3. Ch. 4 deals with the theory of cosmological perturbations and especially, how to obtain the primordial power spectra, a very central object in the analysis. Then, the inflationary model of interest, R^2 -inflation is discussed in Ch. 5 and certain predictions are shown in a qualitative way to gain intuition before the extensive results. Ch. 6 shows how the inflatinary mechanism can be probed with observations today. In Ch. 7 the main tools of the analysis are introduced and the specifications of the considered experiments presented. Finally, Ch. 8 gives the constraints of extended Starobinsky inflation and general cosmological parameters. Ultimately, the content of the thesis takes a turn and explores in the last chapter (9) a deep learning bayesian inference set-up using 21cm tomography. However, not for the previously used inflationary parameters, but dark matter and epoch of reionization parameters.

CHAPTER 2

FLRW Universe

Before diving into the theory of and the motivation for inflation in the next chapter, it is necessary to discuss the fundamental objects used in the description of the Universe. Its evolution is determined by the laws of gravity: General Relativity (GR). This chapter is based on the Baumann TASI Lecture Notes on Inflation [19].

2.1 Geometry

The Universe is on large scales homogeneous and isotropic. This kind of geometry is described by the Friedmann- Lemaître-Robertson-Walker (FLRW) metric $g_{\mu\nu}$, which is defined by the following line element

$$ds^2 = g_{\mu\nu}dx^\mu dx^\nu = -dt^2 + a^2(t) \left[\frac{dr^2}{1 - kr^2} + r^2 d\Omega^2 \right], \quad (2.1)$$

where $a(t)$ is the scale factor and $k \in \{-1, 0, 1\}$ is the curvature. This particular line element has a rescaling symmetry

$$a \rightarrow \lambda a, r \rightarrow r/\lambda, k \rightarrow \lambda^2. \quad (2.2)$$

It is therefore possible to set $a_0 \equiv a(t_0) \equiv 1$, where t_0 equals the time today, by using the rescaling symmetry in a suitable way.

It can be convenient to make the variable transformation $d\chi = dr/\sqrt{1 - kr^2}$ to rewrite the line element to

$$ds^2 = -dt^2 + a^2(t) [d\chi^2 + S_k^2(\chi)d\Omega^2], \quad (2.3)$$

with

$$S_k(\chi) = \begin{cases} \sinh \chi, k = -1 \\ \chi, k = 0 \\ \sin \chi, k = +1 \end{cases}. \quad (2.4)$$

Since the propagation of light depends on the causal structure of the universe, it is sensible to introduce conformal time

$$d\tau = \frac{dt}{a(t)}, \quad (2.5)$$

so that Eq. (2.3) becomes

$$ds^2 = a^2(\tau) [-d\tau^2 + (d\chi^2 + S_k^2(\chi)d\Omega^2)]. \quad (2.6)$$

The metric has factorised into a Minkowskian one and the scale factor. Since light travels on null geodesics $ds^2 = 0$, the propagation of light in FLRW is the same as in Minkowski space if one transforms to conformal time, which is called conformal flatness.

2.2 Kinematics

A particle of mass m follows the trajectory, which extremises proper time $\delta s/c$. The curve is called a geodesic and the extremal path satisfies the geodesic equation

$$\frac{dU^\mu}{ds} + \Gamma_{\alpha\beta}^\mu U^\alpha U^\beta = 0, \quad (2.7)$$

where U^μ is the four-velocity and $\Gamma_{\alpha\beta}^\mu$ are the Christoffel symbols defined by

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

Given the definition of the four-momentum $P^\mu = mU^\mu$ and the FLRW metric, one gets

$$\begin{aligned} P^0 \frac{dP^\mu}{dt} &= -\Gamma_{\alpha\beta}^\mu P^\alpha P^\beta \\ &= -(2\Gamma_{0j}^\mu P^0 + \Gamma_{ij}^\mu) P^j, \end{aligned} \quad (2.9)$$

since every other index combination of the Christoffel symbols vanishes.

There are a few things to notice in this equation. If the particle is at rest in the comoving frame, it will stay at rest. Furthermore, the physical three-momentum $p^2 = g_{ij}P^i P^j = a^2(t)\gamma_{ij}P^i P^j$ of every particle decays with the expansion of the universe, since

$$\frac{\dot{p}}{p} = -\frac{\dot{a}}{a} \Rightarrow p \propto \frac{1}{a}. \quad (2.10)$$

2.3 Redshift

In the quantum mechanical description of the photon its wavelength is given by $\lambda = h/p$. Because of Eq.(2.10) the momentum of the photon evolves as $a^{-1}(t)$. Hence, a photon emitted at time t_1 with wavelength λ_1 is observed at t_0 with the wavelength

$$\lambda_0 = \frac{a(t_0)}{a(t_1)} \lambda_1. \quad (2.11)$$

Since $a(t_0) > a(t_1)$ the wavelength of the photon increases, it is redshifted. The amount of redshift is parametrised by the redshift parameter

$$z \equiv \frac{\lambda_0 - \lambda_1}{\lambda_1}. \quad (2.12)$$

2.4 Dynamics

To understand the dynamics of the universe it is necessary to consider the Einstein equation

$$G_{\mu\nu} = 8\pi GT_{\mu\nu} - \Lambda g_{\mu\nu}, \quad (2.13)$$

where $G_{\mu\nu}$ denotes the Einstein tensor characterising the spacetime curvature, $T_{\mu\nu}$ is the stress-energy tensor and Λ describes the cosmological constant.

The homogeneity and isotropy requirements force the stress-energy tensor to be the one of a perfect fluid,

$$T_{\mu\nu} = (\rho + P)U_\mu U_\nu - P g_{\mu\nu}, \quad (2.14)$$

where ρ is the energy density and P is the pressure of the fluid. The structure of the Einstein tensor forces the covariant conservation equation

$$\nabla_\mu T^\mu{}_\nu = 0 \quad (2.15)$$

and the continuity equation follows

$$\dot{\rho} + 3\frac{\dot{a}}{a}(\rho + P) = 0. \quad (2.16)$$

This equation determines the evolution of the density of the contents of the universe as it expands. Given a constant equation of state $w = P/\rho$ the solutions to Eq.(2.16) scale as

$$\rho \propto a^{-3(1+w)}, \quad (2.17)$$

and therefore

$$\rho \propto \begin{cases} a^{-3}, w = 0, & \text{matter} \\ a^{-4}, w = 1/3, & \text{radiation} \\ a^0, w = -1, & \text{vacuum} \end{cases}. \quad (2.18)$$

In order to relate the matter sources to the evolution of the universe it is necessary to compute the Einstein tensor. It is defined as

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu}, \quad (2.19)$$

where the Ricci tensor is

$$R_{\mu\nu} = \partial_\lambda \Gamma_{\mu\nu}^\lambda - \partial_\nu \Gamma_{\mu\lambda}^\lambda + \Gamma_{\lambda\rho}^\lambda \Gamma_{\mu\nu}^\rho - \Gamma_{\mu\lambda}^\rho \Gamma_{\nu\rho}^\lambda \quad (2.20)$$

and the Ricci scalar is

$$R = g^{\mu\nu} R_{\mu\nu}. \quad (2.21)$$

Given the metric of the FLRW universe these are all computable objects and read as follows

$$R = -6 \left[\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a} \right)^2 + \frac{k}{a^2} \right] \quad (2.22)$$

$$G^0{}_0 = 3 \left[\left(\frac{\dot{a}}{a} \right)^2 + \frac{k}{a^2} \right] \quad (2.23)$$

$$G^i_j = \left[2\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2 + \frac{k}{a^2} \right] \delta_j^i. \quad (2.24)$$

All other components of the Einstein tensor are equal to zero. Combining the stress-energy tensor with the Einstein tensor yields the two Friedmann equations:

$$H^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2} \quad (2.25)$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3P), \quad (2.26)$$

where H is the Hubble parameter, $H \equiv \dot{a}/a$. Note, that the continuity equation (2.16) can be derived from the two Friedmann equations and those are not independent. The first Friedmann equation Eq.(2.25) is often written in terms of dimensionless density parameters

$$\Omega_i \equiv \frac{\rho_i}{\rho_{crit}} \quad (2.27)$$

where $i \in \{r, m, k, \Lambda\}$, denoting radiation, matter, curvature and dark energy, respectively. ρ_{crit} denotes the critical density today, hence

$$\frac{H^2}{H_0^2} = \Omega_r a^{-4} + \Omega_m a^{-3} + \Omega_k a^{-2} + \Omega_\Lambda. \quad (2.28)$$

Due to the different scaling of the various components of the universe, it is most of the time dominated by one single component. At first it was radiation, then matter and dark energy at last. For a universe filled with one single component i the Friedman equation Eq.(2.28) reduces to

$$\frac{\dot{a}}{a} = H_0 \sqrt{\Omega_i} a^{-3(1+w_i)/2}. \quad (2.29)$$

The evolution of the scale factor depends on the dominating component of the universe, as can be seen by the solution to the Friedmann equation:

$$a(t) \propto \begin{cases} t^{1/2}, & \text{radiation dominated} \\ t^{2/3}, & \text{matter dominated} \\ e^{Ht}, & \text{dark energy dominated} \end{cases}. \quad (2.30)$$

CHAPTER 3

Inflation

The standard Big Bang scenario resulting from FLRW cosmology has some major problems. For example, it cannot explain why the Universe is so homogeneous and isotropic on large scales today. Inflation, an early period of exponential expansion, solves this problem. This chapter introduces this problem and its solution in more detail and discusses the necessities to successfully set up an inflationary theory. Like the previous chapter it follows the Baumann TASI lecture notes on inflation [19]. Here, the reduced Planck mass, instead of Newton's gravitational constant, is used:

$$M_P = \sqrt{\frac{\hbar c}{8\pi G}} = 2.4 \cdot 10^{18} \text{ GeV} \quad (3.1)$$

3.1 Horizon Problem

The size of a patch in the sky, which is causally connected, is determined by how far light can travel in a certain amount of time. Using the definition of conformal time in Eq.(2.5) and $ds^2 = 0$ the path of photon in the FLRW universe is

$$\Delta\chi(\tau) = \pm\Delta\tau, \quad (3.2)$$

where the $+$ and $-$ correspond to outgoing and incoming photons, respectively. There are two horizons in cosmology, illustrated in Fig. 3.1:

- Particle horizon: The greatest comoving distance from which an observer can receive a signal is given by

$$\chi_{\text{ph}}(\tau) = \tau - \tau_i = \int_{\tau_i}^{\tau} \frac{dt}{a(t)}. \quad (3.3)$$

χ_{ph} is called the particle horizon. Every causal influence on the observer has to come from within this region.

- Event horizon: The greatest distance from which an observer at time t_f can receive a signal emitted later than t is called the event horizon and is given by

$$\chi_{\text{eh}}(\tau) = \tau_f - \tau = \int_{\tau}^{\tau_f} \frac{dt}{a(t)}. \quad (3.4)$$

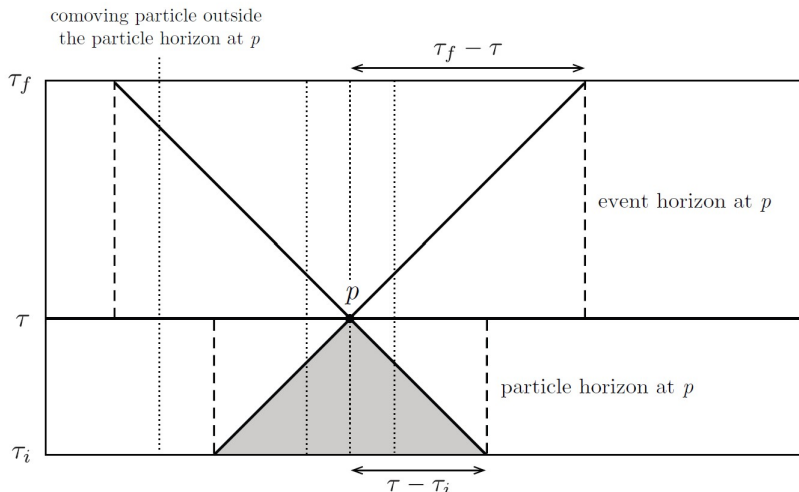


Figure 3.1: Spacetime diagram illustrating the concept of the two horizons, adapted from Ref.[19].

For the horizon problem, the relevant one is the particle horizon and it can be better understood by considering the Hubble sphere. Eq. (3.3) can be rewritten as

$$\chi_{\text{ph}}(a) = \int_{t_i}^t \frac{dt}{a} = \int_{\ln a_i}^{\ln a} (aH)^{-1} d \ln a. \quad (3.5)$$

With a stress-energy tensor describing a perfect fluid with a constant equation of state the particle horizon follows as

$$\chi_{\text{ph}} = \frac{2H_0^{-1}}{1+3w} \left[a^{\frac{1}{2}(1+3w)} - a_i^{\frac{1}{2}(1+3w)} \right], \quad (3.6)$$

where the comoving Hubble radius

$$(aH)^{-1} = H_0^{-1} a^{\frac{1}{2}(1+3w)} \quad (3.7)$$

is used. It can be clearly seen, that the particle horizon gets its largest contribution from late times, if $w > -1/3$, i.e. if the comoving Hubble radius increases as the universe expands.

The problem arises when one considers the cosmic microwave background (CMB). It was formed just after the formation of the first atoms, called recombination. The photons fell out of equilibrium with electrons as there were a lot less electrons around to scatter off. The CMB is discussed in more detail in Sec. 6.1. Now, the problem is that the every point of CMB has the same temperature up to one part in ten thousand. However, if one considers two points in the sky, it is possible to calculate their particle horizons and notices, that if they are separated by more than one degree, there are no points inside both of their horizons. Which poses the question: How can photons have the same properties if they have never been in contact before?

3.2 Solution to the Horizon Problem

The particle horizon is only dominated by late times if the Hubble radius is constantly increasing. Considering a phase of a shrinking Hubble radius in the early

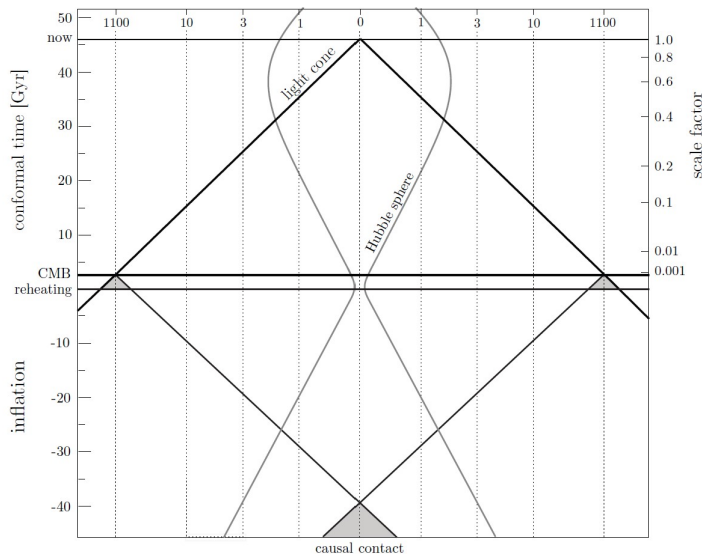


Figure 3.2: Spacetime diagram illustrating the concept of the shrinking Hubble sphere and the causal contact between two points of the CMB, adapted from Ref.[19].

universe,

$$\frac{d}{dt}(aH)^{-1} < 0. \quad (3.8)$$

This pushes the Big Bang singularity to negative conformal time

$$\tau_i = \frac{2H_0^{-1}}{1+3w} a_i^{\frac{1}{2}(1+3w)} \xrightarrow{a_i \rightarrow 0, w < -1/3} -\infty \quad (3.9)$$

A shrinking Hubble sphere requires the equation of state $w < -1/3$ (3.7). For a sufficiently long phase of decreasing Hubble radius all points in the observed CMB are in causal contact and had enough time to thermalise. This solution is illustrated in Fig. 3.2, where the causal contact for different regions of the CMB is made possible by the shrinking Hubble-sphere.

To solve the horizon problem, it is necessary that the observable Universe today fits inside the comoving Hubble sphere at the beginning of inflation,

$$(a_0 H_0)^{-1} < (a_I H_I)^{-1}. \quad (3.10)$$

To estimate the duration of the inflationary period, one can ignore the recent matter- and dark energy-dominated epochs and concentrate on the radiation domination ($H \propto a^{-2}$) and obtain

$$\frac{a_0 H_0}{a_E H_E} \sim \frac{a_E}{a_0} \sim \frac{T_0}{T_E} \sim 10^{-28}. \quad (3.11)$$

For the rough estimation of the growth of the universe during inflation the temperature at the end of inflation is $T_E = 10^{15}$ GeV and $T_0 = 10^3$ eV today. Assuming a constant Hubble parameter during inflation gives

$$\frac{a_E}{a_I} > 10^{28} \text{ or } \ln \frac{a_E}{a_I} > 64. \quad (3.12)$$

Hence, the horizon problem requires around 64 e -folds of inflation.

3.3 Equivalent Ways of Characterising Inflation

Inflation can be defined by the period in which the Hubble sphere is shrinking. However, there are different ways of describing inflation which can be derived from the shrinking Hubble sphere.

- *Accelerated expansion:*

$$\frac{d}{dt}(aH)^{-1} = \frac{d}{dt}(\dot{a})^{-1} = -\frac{\ddot{a}}{(\dot{a})^2} \quad (3.13)$$

Therefore, a shrinking Hubble radius implies an accelerated expansion $\ddot{a} > 0$.

- *Slowly-varying Hubble parameter:*

$$\frac{d}{dt}(aH)^{-1} = -\frac{\dot{a}H + a\dot{H}}{(aH)^2} = -\frac{1}{a}(1 - \epsilon), \quad (3.14)$$

where $\epsilon \equiv -\dot{H}/H^2 < 1$ for an inflationary period.

- *Negative pressure:*

Inserting the continuity equation (2.16) into the ϵ -parameter gives

$$\epsilon_H = -\frac{\dot{H}}{H^2} = \frac{3}{2} \left(1 + \frac{P}{\rho} \right) < 1 \Rightarrow \frac{P}{\rho} = w < -\frac{1}{3}. \quad (3.15)$$

Since the density is always greater than zero inflation occurs if the pressure is negative.

The slowly varying Hubble-parameter motivates a look at the Hubble-function in more detail. During inflation it is slowly varying and can therefore be described sufficiently enough with its Taylor expansion

$$H(\varphi) = \sum_{n=0}^N \frac{1}{n!} \left. \frac{d^n H}{d\varphi^n} \right|_{\varphi_*} (\varphi - \varphi_*)^n, \quad (3.16)$$

where φ_* is a fixed field value, which will get a physical interpretation later on. As it turns out, a convenient way to describe the expansion is with the logarithmic changes [20]

$$\lambda_H^{(n)} = \left(\frac{M_P^2}{4\pi} \right)^n \frac{(H')^{n-1}}{H^n} \frac{d^{n+1} H}{d\varphi^{n+1}} \quad n \geq 1, \quad (3.17)$$

where it is common to give the first 4 parameters each its own name:

$$\lambda^{(1)} = \epsilon_H, \quad \lambda^{(2)} = \eta_H, \quad \lambda^{(3)} = \xi_H^2, \quad \lambda^{(4)} = \omega_H^3. \quad (3.18)$$

3.4 The Physics of Inflation

Having established the need for inflation it is worthwhile to investigate under which conditions it can occur and also stop. A simple model is the scalar *inflaton* field $\varphi(x)$, with the action

$$S_\varphi = \int d^4x \sqrt{-g} \left[\frac{1}{2} g^{\mu\nu} \nabla_\mu \varphi \nabla_\nu \varphi - V(\varphi) \right], \quad (3.19)$$

where g denotes the determinant of the metric. In this model the inflaton is minimally coupled to gravity and has some self-interaction described by $V(\varphi)$. The stress-energy tensor can be found by varying the action with respect to $\delta g^{\mu\nu}$,

$$T_{\mu\nu} = \nabla_\mu\varphi\nabla_\nu\varphi - g_{\mu\nu} \left(\frac{1}{2}g^{\alpha\beta}\nabla_\alpha\varphi\nabla_\beta\varphi - V(\varphi) \right). \quad (3.20)$$

When the stress-energy of this field dominates the universe it controls the evolution. Due to the symmetries of a FLRW universe the field cannot depend on spatial coordinates, hence, it can only depend on t . The energy density and the pressure follow from time-time and space-space components of the stress-energy tensor, respectively:

$$T^0_0 = \rho_\varphi = \frac{1}{2}\dot{\varphi}^2 + V(\varphi) \quad (3.21)$$

$$T^i_j = -P_\varphi\delta^i_j \Rightarrow P_\varphi = \frac{1}{2}\dot{\varphi}^2 - V(\varphi) \quad (3.22)$$

The equation of state is then given by

$$w_\varphi = \frac{P_\varphi}{\rho_\varphi} = \frac{\frac{1}{2}\dot{\varphi}^2 - V(\varphi)}{\frac{1}{2}\dot{\varphi}^2 + V(\varphi)}. \quad (3.23)$$

To fulfil the condition $w < -1/3$ the potential energy of the inflaton has to dominate over the kinetic energy.

Using both Friedman equations (2.25) the time evolution equation of the inflaton field is described by a Klein-Gordon equation

$$\ddot{\varphi} + 3H\dot{\varphi} + \frac{dV}{d\varphi} = 0. \quad (3.24)$$

As seen in the previous section 3.3, inflation only occurs if the Hubble parameter is slowly varying. To describe the situation the first *slow-roll parameter* is introduced

$$\epsilon = -\frac{\dot{H}}{H^2} = \frac{\frac{1}{2}\dot{\varphi}^2}{M_P^2 H}, \quad (3.25)$$

where the inflation condition ($\epsilon < 1$) is only fulfilled if the kinetic energy is small. Since inflation has to be sustained long enough to causally connect all the observables in the CMB the acceleration of the field has to be small as well. This is described by the dimensionless acceleration per Hubble time

$$\delta \equiv -\frac{\ddot{\varphi}}{H\dot{\varphi}}. \quad (3.26)$$

Introducing the second slow-roll parameter

$$\eta \equiv \frac{\dot{\epsilon}}{H\epsilon} = 2(\epsilon - \delta), \quad (3.27)$$

the conditions for an adequately long sustained period of inflation read $\epsilon, |\eta| \ll 1$. Since the slow-roll parameters are small it is possible to simplify the Friedmann equation (2.28) to

$$H^2 \approx \frac{V}{3M_P^2}. \quad (3.28)$$

Hence, the Hubble-parameter is completely determined by the inflaton potential and the equation of motion becomes

$$3H\dot{\varphi} \approx -\frac{dV}{d\varphi} = -V'. \quad (3.29)$$

From these equations it becomes clear that a characterization of slow-roll inflation in terms of the potential is also possible. These are the *potential slow-roll parameters*

$$\epsilon_V \equiv \frac{M_P^2}{2} \left(\frac{V'}{V} \right)^2, \quad |\eta_V| \equiv M_P^2 \frac{|V''|}{V}. \quad (3.30)$$

There is the same condition for successful slow-roll inflation as before, namely $\epsilon_V, |\eta_V| \ll 1$.

The amount of inflation is characterised by the total number of e-folds of accelerated expansion

$$N_{\text{tot}} = \int_{a_{\text{ini}}}^{a_{\text{end}}} d \ln a = \int_{t_{\text{ini}}}^{t_{\text{end}}} H(t) dt, \quad (3.31)$$

where t_{ini} and t_{end} are defined by the boundaries where $\epsilon(t) < 1$. Using Eq. (3.29) and Eq. (3.30) one can write

$$N_{\text{tot}} = \int_{\varphi_{\text{ini}}}^{\varphi_{\text{end}}} \frac{1}{\sqrt{2\epsilon_V}} \frac{|d\varphi|}{M_P}. \quad (3.32)$$

3.5 Reheating

In order to end inflation the potential has to steepen as then the equation of state becomes larger than $-1/3$ and accelerated expansion stops. The inflaton picks up a lot of kinetic energy, since most of the energy density in the universe is in the form of the inflaton potential. Then, it starts to oscillate very rapidly around the minimum of the potential, which is assumed to be approximated by $V(\varphi) = \frac{1}{2}m^2\varphi^2$. The equation of motion becomes

$$\ddot{\varphi} + 3H\dot{\varphi} = -m^2\varphi. \quad (3.33)$$

Since the expansion time scale H^{-1} is much larger than the oscillation period m^{-1} , the friction term $3H\dot{\varphi}$ can be ignored and the inflaton field oscillates with frequency m [19]. The energy continuity equation reads as

$$\dot{\rho}_\varphi + 3H\rho_\varphi = -3HP_\varphi = -\frac{3}{2}H(m^2\varphi^2 - \dot{\varphi}^2). \quad (3.34)$$

Since the average over one oscillation of the r.h.s vanishes the pressure of the field is equal to zero and it behaves like pressureless matter.

The inflaton has to couple to Standard Model (SM) particles, otherwise the universe would be empty. This adds a term to the energy continuity equation,

$$\dot{\rho}_\varphi + 3H\rho_\varphi = -\Gamma_\varphi\rho_\varphi, \quad (3.35)$$

where Γ_φ parametrises the decay rate of the inflaton. After the inflaton has decayed into the usual SM degrees of freedom the standard Hot Big Bang scenario commences.

CHAPTER 4

Cosmological Perturbation Theory

Inflation does not only explain the observed flatness and solves the horizon problem, it also seeds the structures of the Universe through primordial quantum fluctuations, which evolve into macroscopic fluctuations through exponential expansion. In this chapter the formalism of cosmological perturbation theory is discussed and linked to inflation, following the Baumann TASI lecture notes on inflation [19].

4.1 Generalities

To understand how large-scale structures in the universe have formed, it is necessary to consider small perturbations in the early universe. Since the inhomogeneities of the CMB are of order 10^{-5} , it is sufficient to expand all objects of interest as an isotropic average plus some space-dependent perturbation,

$$X(\mathbf{x}, t) = \bar{X}(t) + \delta X(\mathbf{x}, t). \quad (4.1)$$

Perturbations in the FLRW metric and the matter distribution will be coupled to each other through the Einstein equation.

4.1.1 Perturbed Spacetime

To avoid any unnecessary difficulties, it is sensible assume a flat FLRW spacetime, where the metric is defined by

$$ds^2 = a^2(\tau)[-d\tau^2 + \delta_{ij}dx^i dx^j]. \quad (4.2)$$

The perturbation is performed according to Eq. (4.1)

$$g_{\mu\nu} = \bar{g}_{\mu\nu} + \delta g_{\mu\nu}, \quad (4.3)$$

which leads to

$$ds^2 = a^2(\tau)[-(1 + 2A)d\tau^2 - 2B_i dx^i d\tau + (\delta_{ij} + h_{ij})dx^i dx^j], \quad (4.4)$$

where A , B_i and h_{ij} are functions of space and time.

The Einstein equations for scalars, vectors and tensors do not mix. Hence, it is useful to separate the perturbations according to this classification. This procedure is naturally called SVT-decomposition. Any vector can be split into the gradient of a scalar and a divergence-free vector, as can be done for a tensor:

$$B_i = \partial_i B + \hat{B}_i, \quad (4.5)$$

$$h_{ij} = 2C\delta_{ij} + 2\partial_{(i}\partial_{j)}E + 2\partial_{(i}\hat{E}_{j)} + 2\hat{E}_{ij}, \quad (4.6)$$

where

$$\partial_{(i}\partial_{j)}E \equiv (\partial_i\partial_j - \frac{1}{3}\delta_{ij}\nabla^2)E, \quad (4.7)$$

$$\partial_{(i}\hat{E}_{j)} \equiv \frac{1}{2}(\partial_i\hat{E}_j + \partial_j\hat{E}_i). \quad (4.8)$$

The 10 degrees of freedom of the metric thus have been decomposed into 4 + 4 + 2 SVT d.o.f.:

- *scalars*: A, B, C, E
- *vectors*: \hat{B}_i, \hat{E}_i
- *tensors*: \hat{E}_{ij}

The Gauge Problem

These perturbations are not uniquely defined as one can always perform a gauge transformation, e.g. $x^i \mapsto x^i + \xi^i(\tau, \mathbf{x})$. Starting from the unperturbed metric a gauge transformation has the same effect as perturbing the metric itself. Which means in return that perturbations can be gauged away. This clearly imposes a problem, since it is necessary to introduce real, physical perturbations. One way to do so is to look at gauge invariant perturbations, which can be constructed in the following way

$$\begin{aligned} \Psi &\equiv A + \mathcal{H}(B - E') + (B - E')' \\ \hat{\phi} &\equiv \hat{E}'_i - \hat{B}_i \\ \hat{E}_{ij} & \\ \phi &\equiv -C\mathcal{H}(B - E') + \frac{1}{3}\nabla^2 E, \end{aligned} \quad (4.9)$$

where the $X' = \frac{dX}{d\tau}$ and $\mathcal{H} = a'/a$ is the Hubble parameter in conformal time. These gauge invariant perturbations are called the *Bardeen variables*. Another way to solve the problem is to simply fix the gauge. A popular gauge is the *Newtonian gauge*, where $B = E = 0$ implies the metric

$$ds^2 = a^2(\tau)[-(1 + 2\Psi) + (1 - 2\phi)\delta_{ij}dx^i dx^j]. \quad (4.10)$$

Here, the perturbations have been renamed in such a way, that one can see the resemblance to the Bardeen variables.

4.1.2 Perturbed Matter

Similar to the metric, one can also impose small perturbations in the energy-momentum tensor

$$T^\mu_\nu = \bar{T}^\mu_\nu + \delta T^\mu_\nu, \quad (4.11)$$

where

$$\delta T^\mu_\nu = (\delta\rho + \delta P)\bar{U}^\mu\bar{U}_\nu + (\bar{\rho} + \bar{P})(\delta U^\mu\bar{U}_\nu + \bar{U}^\mu\delta U_\nu) - \delta P\delta^\mu_\nu - \Pi^\mu_\nu. \quad (4.12)$$

Redefining the isotropic pressure P allows to find a traceless anisotropic stress tensor Π^μ_ν . In the frame orthogonal to the anisotropic stress tensor ($U^\mu\Pi_{\mu\nu} = 0$), with four-velocity U^μ , one can assume without loss of generality $\Pi^0_0 = 0 = \Pi^0_j$. Since the perturbed four-velocity has to satisfy $g_{\mu\nu}U^\mu U^\nu = -1$, there are only three degrees of freedom, i.e.

$$U_\mu = a(1 + A, -(v_i + B_i)), \quad (4.13)$$

with the variables from Eq.(4.9). Using this, the perturbations in the stress-energy tensor are as follows

$$\begin{aligned} \delta T^0_0 &= \delta\rho \\ \delta T^i_0 &= (\bar{\rho} + \bar{P})v^i \\ \delta T^0_j &= -(\bar{\rho} + \bar{P})(v_j + B_j) \\ \delta T^i_j &= -\delta P\delta^i_j - \Pi^i_j. \end{aligned} \quad (4.14)$$

From now on, q_i will be used as the momentum density $(\bar{\rho} + \bar{P})v_i$, which can be decomposed in a scalar and vector part. The density and matter perturbation have scalar parts only, and the anisotropic stress has scalar, vector and tensor parts. The nomenclature is the same as in the previous SVT-decomposition.

Similar to the space-time perturbations, gauge transformations can absorb the effect of matter perturbations. The result of a coordinate transformation on the stress-energy tensor looks the same as a perturbation. There are various gauge-invariant quantities, for example

$$\bar{\rho}\Delta = \delta\rho + \bar{\rho}'(v + B). \quad (4.15)$$

Δ is called the *comoving-gauge density perturbation*. Alternatively, one can fix the gauge by either setting the scalar momentum density to zero ($q = 0$), which is called *comoving gauge*, or by setting the total density perturbation to zero ($\delta\rho=0$), which is called *uniform density gauge*.

4.1.3 Perturbed Equations

Using the defined metric and matter perturbations, it is possible to work out the effects of the perturbations. It is necessary to choose a gauge, which will be the Newtonian gauge here, where the metric takes the following form

$$g_{\mu\nu} = a^2 \begin{pmatrix} -(1 + 2\Psi) & 0 \\ 0 & (1 - 2\phi)\delta_{ij} \end{pmatrix}, \quad (4.16)$$

and the anisotropic stress is set to zero.

From the perturbed metric one can calculate the perturbed connection. By considering the individual components of the stress-energy conservation equation

$$\nabla_\mu T^\mu_\nu = 0, \quad (4.17)$$

one gets the following perturbed equations:

Continuity Equation

The $\nu = 0$ component of Eq. (4.17) gives the perturbed continuity equation with the zeroth-order and first-order term

$$\bar{\rho}' = -3\mathcal{H}(\bar{\rho} + \bar{P}), \quad (4.18)$$

$$\delta\rho' = -3\mathcal{H}(\delta\rho + \delta P) + 3\phi'(\bar{\rho} + \bar{P}) - \nabla\mathbf{q}, \quad (4.19)$$

respectively. The zeroth-order equation is just the unperturbed continuity equation as one would expect. On the right-hand side of the first-order equation are three terms. The first one corresponds to the dilution of the perturbations due to the background expansion. The second one is a relativistic effect caused by the local expansion rate and the last term accounts for the local fluid flow.

Euler Equation

The $\nu = i$ components of the stress-energy conservation equation (4.17) give the relativistic version of the Euler equation

$$v' + \mathcal{H}v - 3\mathcal{H}\frac{\bar{P}'}{\bar{\rho}'}v = -\frac{\nabla\delta P}{\bar{\rho} + \bar{P}} - \nabla\Psi. \quad (4.20)$$

Einstein Equations

From the perturbed connection one can calculate the Einstein tensor and relate it to the perturbed stress-energy tensor using the Einstein equation. The absence of anisotropic stress implies $\phi = \Psi$, which allows to write all equations in terms of ϕ . Here, only the resulting first order equations will be stated.

From the 00-equation follows

$$\nabla^2\phi = 4\pi Ga^2\bar{\rho}\delta + 3\mathcal{H}(\phi' + \mathcal{H}\phi), \quad (4.21)$$

where $\delta \equiv \delta\rho/\bar{\rho}$. The $0i$ -equation gives

$$\phi' + \mathcal{H}\phi = -4\pi Ga^2(\bar{\rho} + \bar{P})v. \quad (4.22)$$

This results with the 00-equation and the comoving density perturbation in a Poisson equation

$$\nabla^2\phi = 4\pi Ga^2\bar{\rho}\Delta, \quad (4.23)$$

where Δ can be understood as fractional overdensities in the comoving gauge which source the gravitational potential ϕ . Finally, the ii -equation yields at first order

$$\phi'' + 3\mathcal{H}\phi' + (2\mathcal{H}' + \mathcal{H}^2)\phi = 4\pi Ga^2\delta P. \quad (4.24)$$

4.1.4 Conserved Curvature Perturbation

An important gauge invariant quantity is the *comoving curvature perturbation* \mathcal{R} . In Newtonian gauge it is given as

$$\mathcal{R} = -\phi + \mathcal{H}v, \quad (4.25)$$

or in terms of the gravitational potential and its time derivative

$$\mathcal{R} = -\phi - \frac{\mathcal{H}(\phi' + \mathcal{H}\phi)}{4\pi G a^2(\bar{\rho} + \bar{P})}. \quad (4.26)$$

The importance of the comoving curvature perturbations becomes clear by noticing, that they are conserved on super-Hubble scales for adiabatic fluctuations [21].

4.2 Initial Conditions from Inflation

A very important aspect of inflation is that it provides the seed for structure formation in the universe and we do not end up in a completely homogeneous universe. Due to quantum fluctuations of the inflaton field $\delta\varphi(t, \mathbf{x}) = \varphi(t, \mathbf{x}) - \bar{\varphi}(t)$, which can vary spatially, inflation ends at different times at different places in the universe. This can be understood by realising that inflation ends when the field acquires a certain value. This leads to a difference in the amount of expansion that occurs and finally in differences in the local density $\delta\rho(t, \mathbf{x})$.

It is necessary to briefly discuss the definition of a power spectrum and the notation. Starting from some arbitrary perturbations $\delta(\mathbf{x})$ the Fourier transformation is given by

$$\delta(\mathbf{x}) = \int \frac{d^3\mathbf{k}}{(2\pi)^3} \delta(\mathbf{k}) e^{i\mathbf{k}\mathbf{x}} \quad (4.27)$$

and its variance

$$\langle \delta(\mathbf{x}) \rangle^2 = \int \frac{d^3\mathbf{k}}{(2\pi)^3} e^{i\mathbf{k}\mathbf{x}} \int \frac{d^3\mathbf{k}'}{(2\pi)^3} e^{i\mathbf{k}'\mathbf{x}} \langle \delta(\mathbf{k}) \delta^*(\mathbf{k}') \rangle, \quad (4.28)$$

where the correlation is given by

$$\langle \delta(\mathbf{k}) \delta^*(\mathbf{k}') \rangle = (2\pi)^3 \delta_D(\mathbf{k} - \mathbf{k}') P_\delta(\mathbf{k}'), \quad (4.29)$$

with the delta distribution $\delta_D(\mathbf{k} - \mathbf{k}')$ and the dimensionfull power spectrum $P_\delta(\mathbf{k}')$. The delta distribution cancels one integral and one is left with

$$\int \frac{d^3\mathbf{k}}{(2\pi)^3} P_\delta(\mathbf{k}) = \int d \ln \frac{k^3}{(2\pi)^3} P_\delta(\mathbf{k}). \quad (4.30)$$

Now, one can introduce the dimensionless power spectrum $\mathcal{P}_\delta(\mathbf{k})$ as

$$\mathcal{P}_\delta(\mathbf{k}) = \frac{k^3}{(2\pi)^3} P_\delta(\mathbf{k}). \quad (4.31)$$

Note the unfortunately common notation with the marginal difference in P and \mathcal{P} .

4.2.1 From Quantum to Classical

The fluctuation of the inflaton field can be described by quantum harmonic oscillators on small scales, which induce a non-zero variance in the amplitudes

$$\langle |\delta\varphi_k|^2 \rangle \equiv \langle 0 | |\delta\varphi_k|^2 | 0 \rangle \neq 0. \quad (4.32)$$

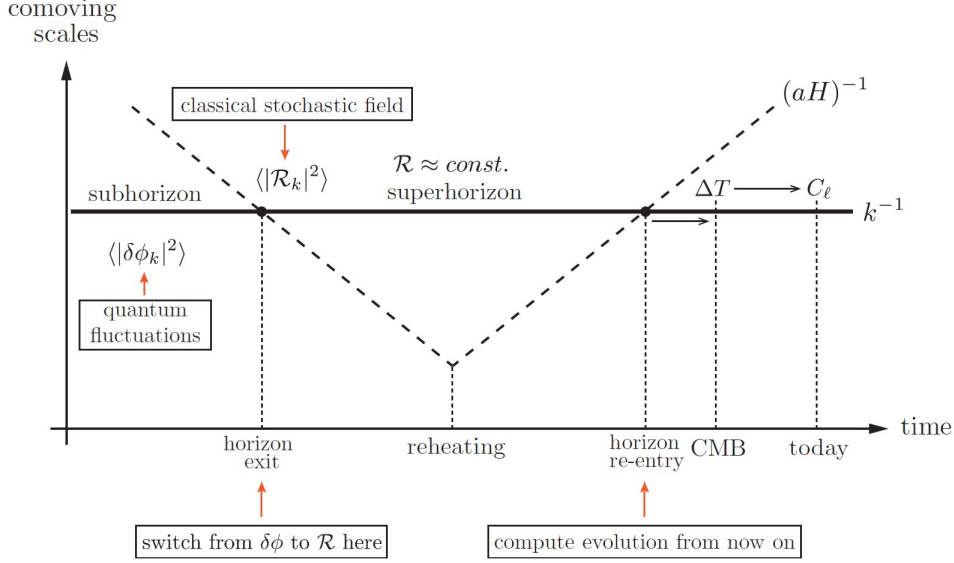


Figure 4.1: Curvature perturbation during and after inflation, adapted from Ref.[19]. ϕ is equivalent to φ in the text.

The wavelength of these fluctuations stays constant in the comoving frame during inflation, but the Hubble radius shrinks. Therefore, these fluctuations become larger than the Hubble radius, called superhorizon, at some point during inflation. At the horizon crossing $k = aH$ it is sensible to describe the fluctuations by the conserved curvature perturbation \mathcal{R} , as they do not evolve outside the horizon. On the superhorizon scale, the nature of the fluctuations is no longer quantum and the expectation value can be understood as the ensemble average of a classical stochastic field. The evolution of perturbations is illustrated in Fig. 4.1. In spatially flat gauge \mathcal{R} is given by

$$\mathcal{R} = -\frac{\mathcal{H}}{\dot{\varphi}}\delta\varphi, \quad (4.33)$$

and its variance

$$\langle |\mathcal{R}_k|^2 \rangle = \left(\frac{\mathcal{H}}{\dot{\varphi}}\right)^2 \langle |\delta\varphi_k|^2 \rangle. \quad (4.34)$$

Most of the fluctuations exit the horizon well before the end of inflation. There are fluctuations on all scales, hence some exit the horizon only right before reheating. However, they are very small scale and not experimentally testable. The fluctuations remain constant during the phase of unknown physics, like reheating. One would not be able to compute the evolution of perturbations in this regime, so it is rather fortunate that the perturbations exit the horizon before this phase and re-enter it afterwards, when it is possible to calculate the evolution equations. It is only therefore, that inflationary theories are testable.

4.2.2 Mukhanov-Sasaki Equation

The action of the inflaton is

$$S = - \int d\tau d^3x \sqrt{-g} \left[\frac{1}{2} g^{\mu\nu} \partial_\mu \varphi \partial_\nu \varphi - V(\varphi) \right]. \quad (4.35)$$

Splitting the field in a background part and its space-time dependent fluctuations $\varphi = \bar{\varphi}(t) + \delta\varphi(x^\mu)$, where its convenient for a moment to change back to normal time, and expanding the action up to second order yields the equation [22]

$$\ddot{Q} + 3H\dot{Q} + \left[\frac{k^2}{a^2} + V_{,\varphi\varphi} - \frac{1}{a^3} \frac{d}{dt} \left(\frac{a^3}{H} \dot{\varphi}^2 \right) \right] Q = 0, \quad (4.36)$$

where $Q = \frac{\dot{\varphi}}{H} \mathcal{R}$ and $V_{,X}$ denotes $\frac{dV}{dX}$. From now on, the bar for the background field will be omitted. This equation is useful to perform numerical calculations later on as it does not discard any terms, except the higher-order ones and is therefore quoted here. However, to perform the quantisation and understand the individual terms it is best to transform back to conformal time and make the following simplifications: The perturbations in the metric are suppressed by the slow-roll parameter ϵ and can be neglected for this discussion. The second-order action reads

$$S^{(2)} = \frac{1}{2} \int d\tau d^3x \left[(f')^2 - (\nabla f)^2 + \left(\frac{a''}{a} - a^2 V_{,\varphi\varphi} \right) f^2 \right], \quad (4.37)$$

where $f = a\delta\varphi$. The potential term can be neglected as $a''/a \gg a^2 V_{,\varphi\varphi}$, giving the equation of motion for the rescaled perturbation of the inflaton field:

$$f'' - \nabla^2 f - \frac{a''}{a} f = 0 \quad (4.38)$$

Fourier transforming this expression yields

$$f''_k + \left(k^2 - \frac{a''}{a} \right) f_k = 0. \quad (4.39)$$

On subhorizon scales $a''/a \ll k^2$ and can be neglected. The resulting equation for each Fourier mode is just the equation of motion of a simple harmonic oscillator

$$f''_k + k^2 f_k = 0. \quad (4.40)$$

4.2.3 Quantum Fluctuations in de Sitter Space

To obtain a realistic model of quantum fluctuation it is necessary to quantise the fields. Therefore, the field $f(\tau, \mathbf{x}) = a\delta\varphi$ and its conjugate momentum $\pi \equiv \frac{\partial \mathcal{L}}{\partial f'} = f'$ have to be promoted to quantum operators with the canonical commutation relation

$$[\hat{f}(\tau, \mathbf{x}), \hat{\pi}(\tau, \mathbf{x}')] = i\delta(\mathbf{x} - \mathbf{x}'). \quad (4.41)$$

The mode expansion is

$$\hat{f}_k(\tau) = f_k(\tau) \hat{a}_k + f_k^*(\tau) \hat{a}_k^\dagger, \quad (4.42)$$

where f_k and its complex conjugate are solutions to the Mukhanov-Sasaki equation (4.36). \hat{a}_k and \hat{a}_k^\dagger are annihilation and creation operators, respectively, with the usual commutation relation.

In a de Sitter universe with $a''/a \approx 2/\tau^2$ the Mukhanov-Sasaki equation has the solution

$$f_k(\tau) = \alpha \frac{e^{-ik\tau}}{\sqrt{2k}} \left(1 - \frac{i}{k\tau} \right) + \beta \frac{e^{ik\tau}}{\sqrt{2k}} \left(1 + \frac{i}{k\tau} \right), \quad (4.43)$$

with constants α, β , which have to be fixed by initial conditions. At very early times (very large negative τ) all modes of interest were inside the Hubble radius $|k\tau| \gg 1$. Therefore, all observable modes had time-independent frequencies

$$k^2 - \frac{2}{\tau^2} \xrightarrow{\tau \rightarrow -\infty} k^2. \quad (4.44)$$

Hence, the Mukhanov-Sasaki equation reduces to the equation for a free field in Minkowski space. This defines the inflationary vacuum, by only considering the minimal excitation state, which corresponds to the positive frequency mode $f_k \propto e^{-ik\tau}$, i.e.

$$\lim_{\tau \rightarrow -\infty} f_k(\tau) = \frac{e^{-ik\tau}}{\sqrt{2k}}, \quad (4.45)$$

which defines the constants of the general solution. The mode function is now given by

$$f_k(\tau) = \frac{e^{-ik\tau}}{\sqrt{2k}} \left(1 - \frac{i}{k\tau} \right). \quad (4.46)$$

The evolution of each mode is completely determined and can be calculated. The field operator is given by

$$\hat{f}(\tau, \mathbf{x}) = \int \frac{d^3k}{(2\pi)^{3/2}} \left[f_k(\tau) \hat{a}_k + f_k^* a_k^\dagger \right] e^{i\mathbf{k}\cdot\mathbf{x}} \quad (4.47)$$

and its expectation value vanishes but the variance does not

$$\langle |\hat{f}|^2 \rangle = \int d \ln k \frac{k^3}{2\pi^2} |f_k(\tau)|^2. \quad (4.48)$$

This leads to the definition of the *power spectrum*

$$\mathcal{P}_f(k, \tau) \equiv \frac{k^3}{2\pi^2} |f_k(\tau)|^2. \quad (4.49)$$

Finally, the power spectrum for the inflaton perturbations reads as

$$\mathcal{P}_{\delta\varphi}(k, \tau) = \frac{\mathcal{P}_f(k, \tau)}{a^2} = \left(\frac{H}{2\pi} \right)^2 \left(1 + \left(\frac{k}{aH} \right)^2 \right) \xrightarrow{\text{superhorizon}} \left(\frac{H}{2\pi} \right)^2. \quad (4.50)$$

This shows that the power spectrum is constant on superhorizon scales.

4.2.4 Primordial Perturbations from Inflation

As discussed before, it is sensible to switch from the inflaton perturbations to the conserved curvature perturbations at the horizon crossing. These two are related via Eq. (4.33), which implies for the power spectra

$$\mathcal{P}_{\mathcal{R}} = \frac{1}{2\epsilon} \frac{\mathcal{P}_{\delta\varphi}}{M_P^2} \quad (4.51)$$

with the slow-roll parameter ϵ . Substituting the $\delta\varphi$ power spectrum

$$\mathcal{P}_{\mathcal{R}} = \frac{1}{8\pi^2\epsilon} \frac{H^2}{M_P^2} \Big|_{k=aH} \quad (4.52)$$

evaluated at horizon crossing, where it stays unchanged until horizon re-entry (4.50). If the power spectrum is k -independent, it is called *scale-invariant*. However, since H and ϵ are slowly-varying with time it is expected that $\mathcal{P}_{\mathcal{R}}$ is not perfectly scale-invariant. The deviation can be parametrised by the *scalar spectral index*

$$n_s - 1 \equiv \left. \frac{d \ln \mathcal{P}_{\mathcal{R}}}{d \ln k} \right|_{k=k_*}, \quad (4.53)$$

where k_* is some reference scale, where the spectrum takes a power-law form

$$\mathcal{P}_{\mathcal{R}} = A_s \left(\frac{k}{k_*} \right)^{n_s - 1}, \quad (4.54)$$

with the amplitude A_s . To first order in the slow-roll parameters the spectral index can be expressed as

$$n_s - 1 = -2\epsilon - \eta. \quad (4.55)$$

4.2.5 Gravitational Waves

Tensor perturbations in the spatial metric lead to a spectrum of primordial gravitational waves. The perturbation is given by

$$ds^2 = a^2(\tau)[-d\tau^2 + (\delta_{ij} + 2\hat{E}_{ij})dx^i dx^j]. \quad (4.56)$$

Substituting the perturbations into the Einstein-Hilbert action and expanding to second order gives

$$S^{(2)} = \frac{M_P^2}{8} \int d\tau d^3x a^2 [(\hat{E}'_{ij})^2 - (\nabla \hat{E}_{ij})^2]. \quad (4.57)$$

It is possible to define

$$\frac{M_P}{2} a \hat{E}_{ij} = \begin{pmatrix} f_+ & f_\times & 0 \\ f_\times & -f_+ & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad (4.58)$$

such that the action becomes

$$S^{(2)} = \frac{1}{2} \sum_{I=+, \times} \int d\tau d^3x \left[(f'_I)^2 - (\nabla f_I)^2 + \frac{a''}{a} f_I^2 \right], \quad (4.59)$$

which is just two copies of the action for scalar perturbations (4.37). Therefore, the power spectrum \mathcal{P}_t can be directly inferred from the previous result for \mathcal{P}_f ,

$$\mathcal{P}_t = 2\mathcal{P}_{\hat{E}} = 2 \left(\frac{2}{aM_P} \right)^2 \mathcal{P}_f \quad (4.60)$$

$$\Rightarrow \mathcal{P}_t(k) = \left. \frac{2}{\pi^2} \frac{H^2}{M_P^2} \right|_{k=aH}. \quad (4.61)$$

The tensor amplitude only depends on H . It is therefore a direct measure of the Hubble-parameter during inflation. The scale dependence is defined in analogy to the previous case,

$$\mathcal{P}_t = A_t \left(\frac{k}{k_*} \right)^{n_t}. \quad (4.62)$$

Note, that historically there is no additional 1 in the exponent. Scale-invariance implies now $n_t = 0$. One observable experiments typically look at is the *scalar-to-tensor ratio*

$$r \equiv \frac{A_t}{A_s}. \quad (4.63)$$

CHAPTER 5

Starobinsky Model of Inflation and its Extension

There exists a huge amount of models describing inflation, yet this thesis focuses on one of the best fitting to data [5]: The Starobinsky or R^2 -inflation, but with an additional extension to higher order terms.

5.1 Formalism

The Einstein-Hilbert action of General relativity is given in terms of the Ricci scalar R

$$S_{GR} = \frac{M_P^2}{2} \int d^4x \sqrt{-g} R, \quad (5.1)$$

where g denotes the determinant of the metric, the mostly plus metric convention $(-, +, +, +)$ and $M_P = (8\pi G)^{-1/2}$. The cosmological constant is neglected, as it does not influence the evolution of the universe during this epoch, see Eq. (2.30). Eq. (5.1) can be extended to the so called $f(R)$ -gravity [23, 24] with

$$S_J = \frac{1}{2} \int d^4x \sqrt{-g_J} f(R), \quad (5.2)$$

where $f(R)$ can be in general an arbitrary function of R . Here it will be

$$f(R) = M_P^2 \left(R + \frac{1}{6M^2} R^2 + \frac{c}{36M^4} R^3 \right), \quad (5.3)$$

with free parameters M and c . The action is dimensionless, which demands that $f(R)$ has mass dimension four. Knowing, that the Ricci-scalar has mass dimension two, the dimensionality of M and c follows. Namely, M has mass dimension one and c is dimensionless. For $c = 0$ one recovers the pure Starobinsky (or R^2) action. The subscript J denotes the Jordan frame and E will represent the Einstein frame. The difference between these two frames is, that in the Jordan frame, one can have non-minimal coupling between fields and gravity, while in the Einstein frame only minimal couplings exist. Therefore, the complicated structure of $f(R)$ can be absorbed in an additional field with normal GR. Going from the Jordan frame to

the Einstein frame consists of multiple steps, where the first one is to introduce a scalar field s by a Legendre-transformation

$$S_J = \frac{1}{2} \int d^4x \sqrt{-g_J} [f(s) - f'(s)(R - s)]. \quad (5.4)$$

The prime denotes the derivative with respect to the field s . Varying the action with respect to s yields the relation $R - s = 0$. For the Legendre transform to be well defined, $f(R)$ has to be a convex function, i.e. $f''(R) > 0$ for all R . Therefore, it is only well defined under the following condition

$$f''(R) = M_P^2 \left(\frac{1}{3M^2} + \frac{c}{6M^4} R \right) > 0 \quad (5.5)$$

$$\Rightarrow R = s = \begin{cases} > -\frac{M^2}{2c}, c > 0 \\ < -\frac{M^2}{2c}, c < 0 \end{cases}. \quad (5.6)$$

It is convenient to define the action in the Jordan frame as

$$S_J \equiv \int d^4x \sqrt{-g_J} \left[\frac{M_P^2}{2} \Omega^2 R - V(s) \right], \quad (5.7)$$

with

$$\Omega^2 = \frac{f'(s)}{M_P^2} = 1 + \frac{1}{3M^2} s + \frac{c}{12M^4} s^2 \text{ and } V(s) = \frac{1}{2} [s f'(s) - f(s)]. \quad (5.8)$$

To get from the Jordan frame to the Einstein frame one has to perform a Weyl transformation

$$g_{\mu\nu,E} = \Omega^2 g_{\mu\nu,J}. \quad (5.9)$$

From the transformation property of the metric one can derive the transformed Christoffel symbols, then the Ricci tensor and finally the Ricci scalar, which is the relevant quantity here. The transformation property will only be stated in $d = 4$ dimensions as [25]

$$R_E = \frac{1}{\Omega^2} \left[R_J - 6 \frac{\square_J \Omega}{\Omega} \right], \quad (5.10)$$

$$R_J = \Omega^2 \left[R_E + 6 \frac{\square_E \Omega}{\Omega} - 12 g_E^{\mu\nu} \frac{\Omega_{,\nu} \Omega_{,\nu}}{\Omega^2} \right], \quad (5.11)$$

where $\Omega_{,\mu} \equiv \partial_\mu \Omega$ and $\square_{E,J} \equiv g_{E,J}^{\mu\nu} \partial_\mu \partial_\nu$. The Weyl transformation can be related to a scalar field φ through

$$\varphi \equiv \sqrt{\frac{3}{2}} M_P \ln \Omega^2. \quad (5.12)$$

Inserting this in Eq. (5.11) gives the Ricci scalar in the Jordan frame in terms of the new scalar field and the Ricci scalar in the Einstein frame,

$$R_J = \Omega^2 \left[R_E + \frac{\sqrt{6}}{M_P} g_E^{\mu\nu} \partial_\mu \partial_\nu \varphi - \frac{1}{M_P^2} g_E^{\mu\nu} \partial_\mu \varphi \partial_\nu \varphi \right], \quad (5.13)$$

which can then be inserted in Eq. (5.7), yielding the action in the Einstein frame

$$\begin{aligned} S_E &= \int d^4x \frac{\sqrt{-g_E}}{\Omega^4} \left[\frac{M_P^2}{2} \Omega^4 \left(R_E + \frac{\sqrt{6}}{M_P} g_E^{\mu\nu} \partial_\mu \partial_\nu \varphi - \frac{1}{M_P^2} g_E^{\mu\nu} \partial_\mu \varphi \partial_\nu \varphi \right) - V(s) \right] \\ &= \int d^4x \sqrt{-g_E} \left[\frac{M_P^2}{2} R_E - \frac{1}{2} g_E^{\mu\nu} \partial_\mu \varphi \partial_\nu \varphi - V_E(\varphi) \right], \end{aligned} \quad (5.14)$$

where $V_E(\varphi) = V(s(\varphi))/\Omega^4$. The surface term $\propto g_E^{\mu\nu} \partial_\mu \partial_\nu \varphi$ vanishes. To find the potential in terms of φ one needs the relations between s and φ . To recover it, the following equation is inverted

$$e^{\sqrt{\frac{2}{3}} \frac{\varphi}{M_P}} = 1 + \frac{1}{3M^2} s + \frac{c}{12M^4} s^2 \quad (5.15)$$

$$\Rightarrow s(\varphi) = \begin{cases} \frac{2M^2}{c} \left[\pm \sqrt{1 + 3c(e^{\sqrt{\frac{2}{3}} \frac{\varphi}{M_P}} - 1) - 1} - 1 \right] & \text{for } c \neq 0 \\ -3M^2 \left[1 - e^{\sqrt{\frac{2}{3}} \frac{\varphi}{M_P}} \right] & \text{for } c = 0 \end{cases}. \quad (5.16)$$

Comparing the formula for $s(\varphi)$ with the convexity condition for the Legendre transform in Eq. (5.6) shows, that one has to take the solution with the minus sign for $c < 0$ and the one with the plus sign for $c > 0$. Now, the potential can be expressed in terms of the field φ

$$V_E(\varphi) = \frac{V(s(\varphi))}{\Omega^2} = \frac{M_P^2 \left[\frac{cs(\varphi)^3}{M^2} + 3s(\varphi)^2 \right]}{36M^2 \left[1 + \frac{s(\varphi)}{3M^2} + \frac{cs(\varphi)^2}{12M^4} \right]^2}. \quad (5.17)$$

However, for $c < 0$ the potential does not have the needed properties for inflation, as it is unbounded from below. Hence, one can exclude $c < 0$ as a possible extension of Starobinsky inflation. For $c = 0$ the usual Starobinsky potential is obtained

$$V_E(\varphi) = \frac{3M_P^2 M^2}{4} \left(1 - e^{-\sqrt{\frac{2}{3}} \frac{\varphi}{M_P}} \right)^2. \quad (5.18)$$

Comparing this action to the action of single-field inflation (3.19) shows, that the formalism of Ch. 3 is applicable. Effectively, Starobinsky inflation is equivalent to single-field inflation with a particular potential.

5.2 Observables in Starobinsky Inflation

For now, c will be set to zero and the well-known Starobinsky inflation is investigated. In the slow-roll approximation the potential slow-roll parameters are given as

$$\epsilon_V = \frac{M_P^2}{2} \left(\frac{V_{E,\varphi}}{V_E} \right)^2, \quad \eta_V = M_P^2 \frac{V_{E,\varphi\varphi}}{V_E}. \quad (5.19)$$

The derivatives of the potential with respect to the field are given by

$$V_{E,\varphi} = \sqrt{\frac{3}{2}} M_P M^2 e^{-\sqrt{\frac{2}{3}} \frac{\varphi}{M_P}} \left(1 - e^{-\sqrt{\frac{2}{3}} \frac{\varphi}{M_P}} \right) \quad (5.20)$$

$$V_{E,\varphi\varphi} = M^2 e^{-\sqrt{\frac{2}{3}} \frac{\varphi}{M_P}} \left(2e^{-\sqrt{\frac{2}{3}} \frac{\varphi}{M_P}} - 1 \right). \quad (5.21)$$

Therefore, the potential slow-roll parameters are

$$\epsilon_V = \frac{4}{3} \left(e^{\sqrt{\frac{2}{3}} \frac{\varphi}{M_P}} - 1 \right)^{-2} \quad (5.22)$$

$$\eta_V = \frac{4}{3} \left[2 + 3 \left(e^{-\sqrt{\frac{2}{3}} \frac{\varphi}{M_P}} - 1 \right)^{-1} + \left(e^{-\sqrt{\frac{2}{3}} \frac{\varphi}{M_P}} - 1 \right)^{-2} \right]. \quad (5.23)$$

The observables are the spectral index n_s , the scalar amplitude A_s and the tensor-to-scalar ratio r , which can be written in terms of the potential slow-roll parameters

$$n_s = 1 + 2\eta_V - 6\epsilon_V \quad (5.24)$$

$$A_s = \frac{V_E}{24\pi^2 M_P^4 \epsilon_V} \quad (5.25)$$

$$r = \frac{A_t}{A_s} \quad \text{with} \quad A_t = \frac{2V_E}{3\pi^2 M_P^4}. \quad (5.26)$$

The observational constraints from Planck 2018 on these parameters are [5]

$$A_s^{\text{Pl}} = (2.099 \pm 0.014) \cdot 10^{-9} \quad (5.27)$$

$$n_s^{\text{Pl}} = 0.9649 \pm 0.0042 \quad (5.28)$$

$$r^{\text{Pl}} < 0.056. \quad (5.29)$$

The potential slow-roll parameters have to be evaluated at the field value φ_* , which corresponds to the value, where the pivot scale exits the horizon. This happened about 60 e -foldings before the end of inflation and can therefore be calculated with Eq. (3.32) yielding

$$60 = N \approx \int_{\varphi_{\text{end}}}^{\varphi_*} \frac{1}{\sqrt{2\epsilon_V}} \frac{1}{M_P} d\varphi. \quad (5.30)$$

In order to find φ_* the inflaton value at the end of inflation φ_{end} is needed. It is defined by the ending of slow-roll, i.e

$$\epsilon_V(\varphi_{\text{end}}) = 1 \quad (5.31)$$

$$\Rightarrow \varphi_{\text{end}} = M_P \sqrt{\frac{3}{2}} \ln \left(\frac{2}{\sqrt{3}} + 1 \right). \quad (5.32)$$

Solving the integral of Eq. (5.30) yields:

$$\begin{aligned} 60 &= \frac{1}{M_P} \sqrt{\frac{3}{8}} \int_{\varphi_{\text{end}}}^{\varphi_*} \left(e^{\sqrt{\frac{2}{3}} \frac{\varphi}{M_P}} - 1 \right) d\varphi \\ &= -\frac{3}{4} \left[\frac{2}{\sqrt{3}} + 1 - \ln \left(\frac{2}{\sqrt{3}} + 1 \right) \right] + \frac{3}{4} e^{\sqrt{\frac{2}{3}} \frac{\varphi_*}{M_P}} - \varphi_* \end{aligned} \quad (5.33)$$

Since this equation cannot be solved analytically it has to be solved numerically. The solution is

$$\varphi_* \approx 5.493M_P. \quad (5.34)$$

For the inflationary model to be successful, it has to match with the observations from Planck. The only free parameter of the theory M can be calculated from the scalar amplitude A_s (5.26), as this the only observable, which depends on it. For Starobinsky inflation the value reads

$$M = 1.084 \cdot 10^{-5} M_P \quad (5.35)$$

and the prediction for the other two observables is

$$n_s = 0.9689 \quad (5.36)$$

$$r = 0.00277, \quad (5.37)$$

which is consistent with the data.

5.3 Field Dynamics in Starobinsky Inflation

The previous section dealt with Starobinsky Inflation in the slow-roll approximation. However, exact results can be obtained by solving the background equation. This leads to the time evolution of the inflaton field.

The background equation is

$$\ddot{\varphi} + 3H\dot{\varphi} + \frac{dV}{d\varphi} = 0, \quad (5.38)$$

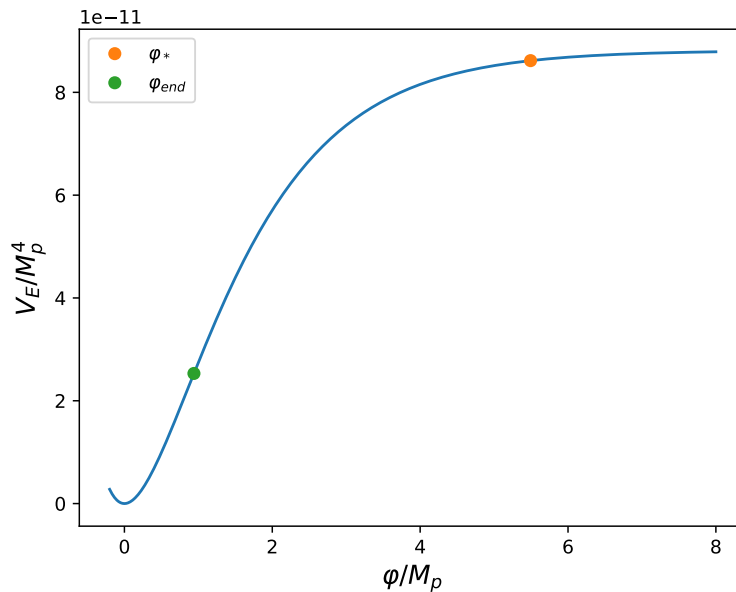


Figure 5.1: Illustration of the inflation potential with the field values, where the pivot scale exits the horizon φ_* and when inflation ends φ_{end} .

where the Hubble function is given by

$$H^2 = \frac{1}{3M_P^2} \left[\frac{1}{2}\dot{\varphi}^2 + V \right]. \quad (5.39)$$

V denotes the potential defined in Eq. (5.18). The Hubble function can be written in terms of the scale factor as

$$H = \frac{\dot{a}}{a} = \frac{d}{dt} \ln a. \quad (5.40)$$

Thus, to obtain the field evolution the following system of coupled differential equations has to be solved:

$$\begin{aligned} (1) \quad & \ddot{\varphi} + \sqrt{\frac{1}{3M_P^2} \left[\frac{1}{2}\dot{\varphi}^2 + \frac{3M^2M_P^2}{4} \left(1 - e^{-\sqrt{\frac{2}{3}}\frac{\varphi}{M_P}} \right)^2 \right]} \dot{\varphi} + \sqrt{\frac{3}{2}}M_P M^2 e^{-\sqrt{\frac{2}{3}}\frac{\varphi}{M_P}} \left(1 - e^{-\sqrt{\frac{2}{3}}\frac{\varphi}{M_P}} \right) = 0 \\ (2) \quad & \frac{d}{dt} \ln a - \sqrt{\frac{1}{3M_P^2} \left[\frac{1}{2}\dot{\varphi}^2 + \frac{3M_P^2M^2}{4} \left(1 - e^{-\sqrt{\frac{2}{3}}\frac{\varphi}{M_P}} \right)^2 \right]} = 0 \end{aligned} \quad (5.41)$$

Choosing the right boundary conditions is crucial as this has a large impact on the solution. Three initial values have been chosen at the initial time t_{ini} . Namely,

$$\begin{aligned} \varphi(t_{ini}) &= \varphi_* \\ \dot{\varphi}(t_{ini}) &= 0 \\ \ln a(t_{ini}) &= 0, \end{aligned} \quad (5.42)$$

where φ_* is the value obtained in the previous section (5.34). It denotes the field value, where 60 e -folds of inflation are still to occur. However, this value was obtained in the slow-roll approximation and is therefore only an initial guess. The time derivative of the inflaton field has to be zero, which can be clearly seen from the form of the potential in Fig. 5.1. Given the solution for $\varphi(t)$ the point in time when inflation ends can be determined with the condition

$$1 = \epsilon(t_{end}) = \frac{\dot{\varphi}^2(t_{end})}{2M_P^2 H(t_{end})}. \quad (5.43)$$

The amount of e -folds that happened during inflation are given by the difference of the logarithmic scale factor at the initial and final time,

$$N = \ln a(t_{end}) - \ln a(t_{ini}) \approx 50 - 60. \quad (5.44)$$

Setting $N = 60$ and varying the initial field value to obtain this N gives $\varphi(t_{ini}) = 5.421$, which is slightly lower than φ_* due to the previously made slow-roll approximation.

The time evolution of interesting quantities can be seen in Fig. 5.2. They are rescaled, such one can see their behaviour at the same time. From the logarithm of the scale factor a , one can directly see how many e -folds of inflation occurred, as $\log(a(t_{end})) - \log(a(t_{ini})) \approx 60$. The oscillatory behaviour in many of the parameters can be explained with the field, rolling down into the minimum of the potential and then starting to oscillate. If one would include the decay of the inflaton to other particles, this behaviour would not be visible, since it would decay before reaching the minimum. The slow-roll phase of the universe is also clearly visible in this figure, where Hubble function is only slowly changing

5.4 Primordial Power Spectrum

So far, only the background evolution of the inflaton field was considered, but fluctuations around this value are necessary to explain the structure in the universe. The fluctuations are best described by the previously introduced gauge invariant Mukhanov-Sasaki variable Q . The time evolution is given by

$$\ddot{Q} + 3H\dot{Q} + \left[\frac{k^2}{a^2} + V'' - \frac{1}{a^3} \frac{d}{dt} \left(\frac{a^3}{H} \dot{\varphi}^2 \right) \right] Q = 0, \quad (5.45)$$

where k is the wavenumber, the equation is motivated in Sec. 4.2.2. To solve this differential equation one chooses the commonly used Bunch-Davies initial conditions. In the end, the quantity of interest is the primordial power spectrum of the curvature perturbations \mathcal{R} , which are related to the Mukhanov-Sasaki variable by

$$\mathcal{R} = \frac{H}{\dot{\varphi}} Q. \quad (5.46)$$

Finally, the power spectrum reads as

$$\mathcal{P}_{\mathcal{R}}(t, k) = \frac{k^3}{2\pi} |\mathcal{R}|^2. \quad (5.47)$$

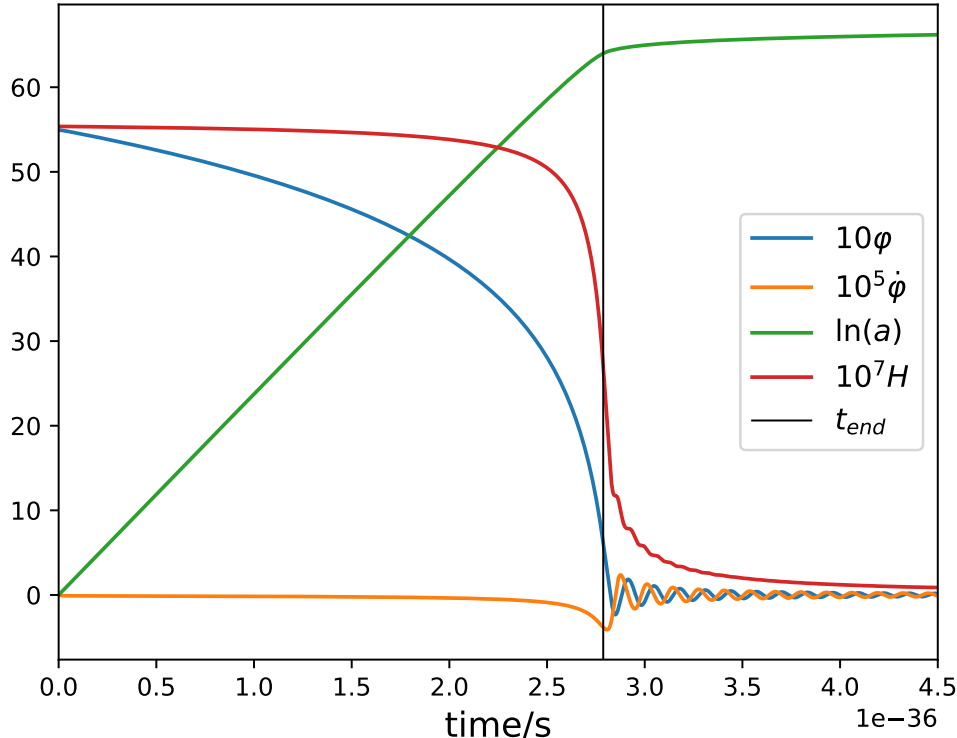


Figure 5.2: Background evolution of various quantities as function of time, rescaled to show the connection. With the inflaton field φ in blue, its derivative $\dot{\varphi}$ in orange, the Hubble function H in red and the logarithm of scale factor showing the e -folds of inflation in green. The end of inflation is denoted by t_{end} .

For a fixed k the time evolution of the curvature perturbation is shown in Fig. 5.3. Curvature perturbations once they exit the horizon, which can be very clearly seen in the plot. The Horizon exit occurs when $k = aH$. In order to get to the primordial power spectrum as a function, which only depends on k , one evaluates $\mathcal{P}_{\mathcal{R}}$ at the Horizon exit. The primordial power spectrum in the Starobinsky Inflation scenario is shown in Fig. 5.4a for multiple values of M to illustrate its effect. The power spectra are computed with CLASS (see Sec. 7.6).

5.5 R^3 -Extension

The most natural way to extend the Starobinsky model of inflation is to simply add a R^3 -term with an additional dimension-free parameter. The effective nature of the Lagrangian becomes even more visible now. One looks at gravity from a phenomenological point of view, where the effects from an extension of GR are absorbed in the expansion of the Lagrangian around the Ricci-scalar. The presence of this term may hint towards the UV-completion of Einstein gravity, as the higher order curvature correspond the quantum effects. Clearly, the question is now, which effect does this new term have on the known observables?

The modification to the potential for different values of the parameter c can be seen in Fig. 5.5. For positive field values, it is only monotonically rising if $c = 0$. For $c > 0$ the potential has a maximum. If the initial field value would be larger than corresponding value inflation would not happen. In order to get the necessary amount of inflation the field has to stay in the slow-roll regime ($V \ll \dot{\phi}$) for a sufficiently long time. If c is too large this condition might not be fulfilled because the potential is not flat enough in the region around the maximum.

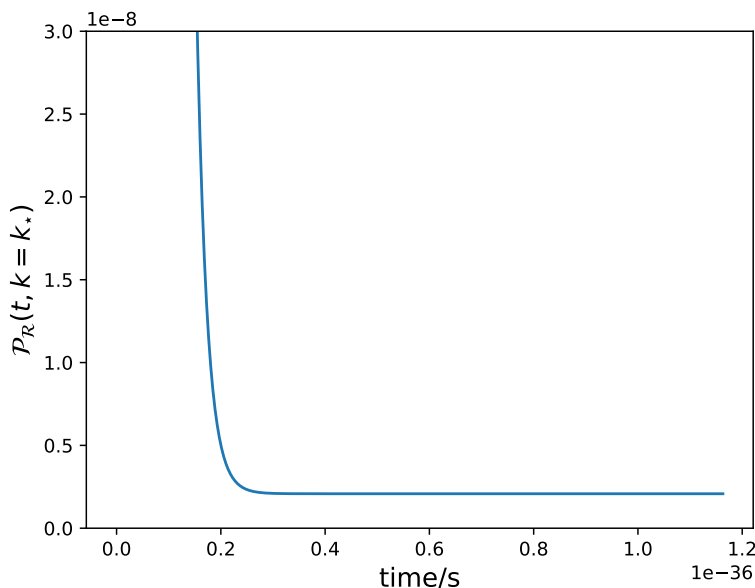
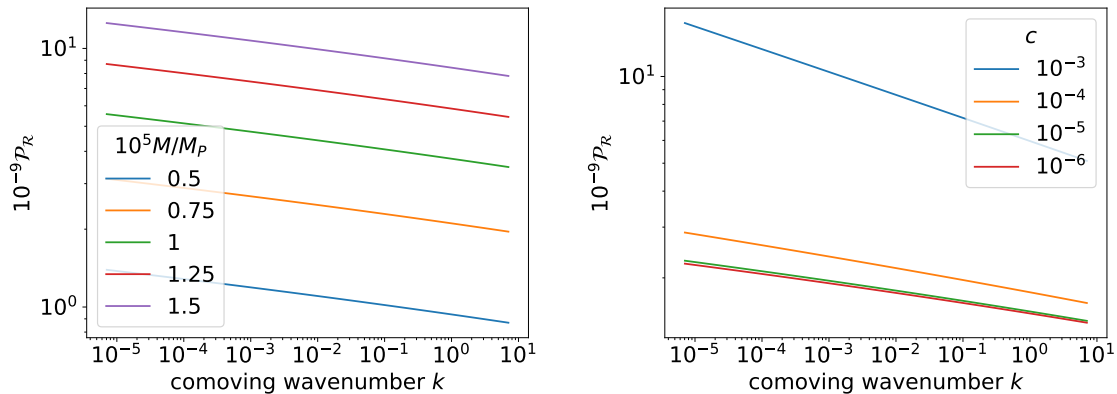


Figure 5.3: Evolution of the curvature perturbation for the reference mode $k = 0.05 \text{ Mpc}^{-1}$.



(a) Pure Starobinsky ($c = 0$), but different values of M . (b) Fixed $M = 1.2 \cdot 10^{-5} M_P$ but varying c .

Figure 5.4: Primordial power spectrum for different values of M and c

To quantify this qualitative argument the time evolution of the background field can be done in exactly the same way as in pure Starobinsky inflation. Assuming that the initial field value is at the maximum of the potential one can get which value of c is needed to get at least a certain amount of e -folds. For this calculation the other parameter of the model M is fixed to one value. The possible amount of e -folds is not influenced by M , it only modifies the overall scale while leaving the shape unchanged. Demanding a certain number of e -folds restricts the possible parameter values for c already quite drastically.

In order to cross-check the upcoming numerical results from the Bayesian inference analysis it is sensible to study the effect of the new parameter on the observables A_s and n_s in the slow-roll approximation. The dependence of these observables on c can be seen in Fig. 5.7 and the effect of c on the full primordial power spectrum

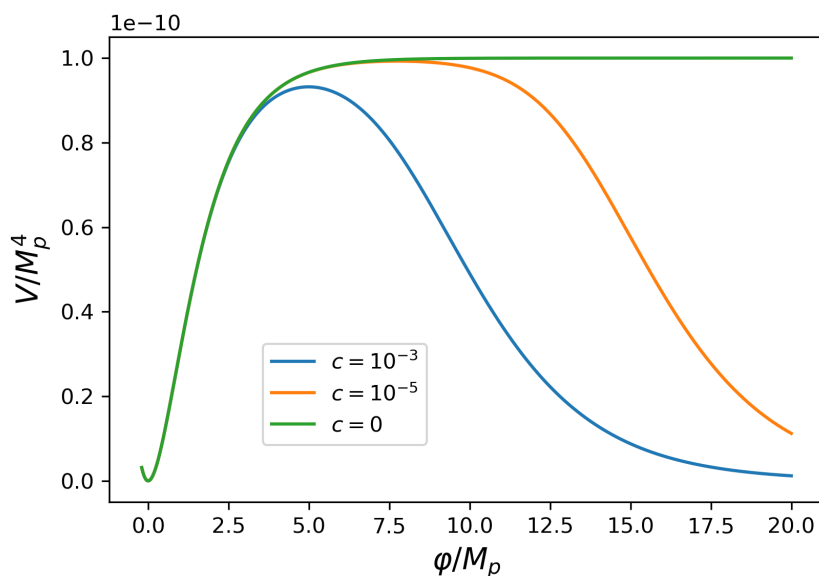


Figure 5.5: Illustration of the inflationary potential for different values of c .

is shown in Fig. 5.4b. Since R^3 is merely an extension of the successful Starobinsky inflation, $c = 0$ has to be compatible with the observations. For large values of c the resulting values of A_s and n_s are incompatible with Planck data.

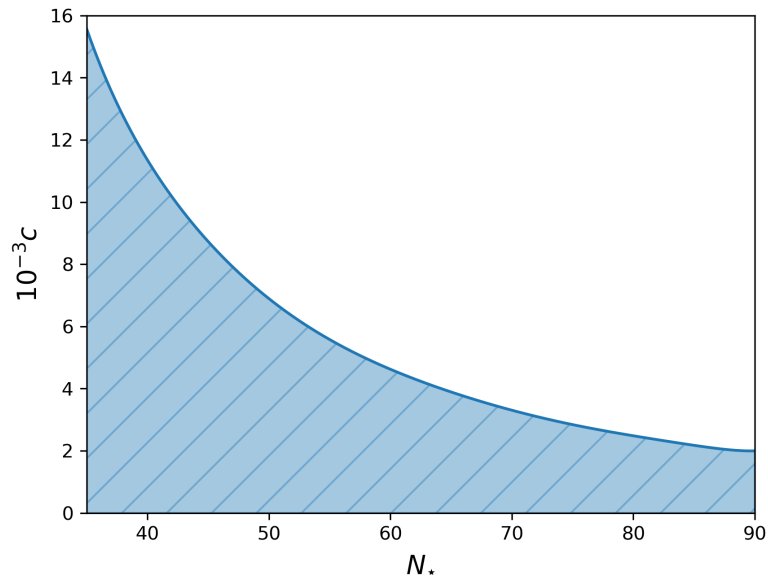


Figure 5.6: Allowed values for c to obtain at least N_* e -folds of inflation

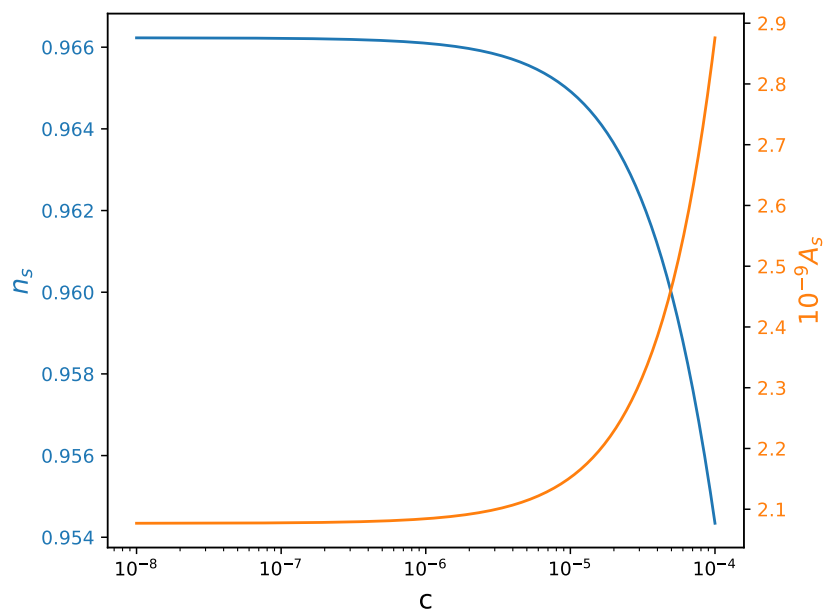


Figure 5.7: Slow-roll prediction for n_s and A_s as a function of c for fixed M

CHAPTER 6

Contact with Observations

After the perturbations exit the horizon they stay constant, as discussed in the previous section (Sec. 5.4), but at some point they re-enter and have to be time evolved. The effect of the primordial fluctuations can then be seen at later times in the anisotropies of the CMB as well as in the matter density fluctuations, observable through the distribution of galaxy and large scale structure. The time evolution is done with the Boltzmann-Code CLASS (cosmic linear anisotropy solving system, Sec. 7.6).

6.1 Cosmic Microwave Background

In the early Universe, there is a thermal ensemble of photons. For photons to be in a thermal equilibrium they need to couple an electromagnetically charged heat bath, as they do not interact with each other. In the early Universe, photons scatter off free electrons and nuclei until the temperature drops to a point, where the first atoms are formed. Now, the photons decouple from the matter content and propagate along straight lines. Since, this happens almost instantly, we can observe the photons coming from a spherical surface, the so called *last scattering surface*. In this work, only a brief introduction to the CMB power spectrum and polarisation is presented, for a more detailed overview see Ref.[26], on which this introduction is based.

6.1.1 Statistics of Temperature Anisotropies

The curvature perturbations \mathcal{R} inside the horizon lead to density fluctuations $\delta\rho$ in the primordial plasma which are imprinted in temperature fluctuations $\Delta T(\mathbf{n})$ of the CMB, where \mathbf{n} denotes the direction in the Sky. The background temperature is $T_0 = 2.726\text{K}$. Due to the spherical nature of the observation it is sensible to perform a harmonic expansion

$$\Theta(\mathbf{n}) \equiv \frac{\Delta T(\mathbf{n})}{T_0} = \sum_{\ell,m} a_{\ell m} Y_{\ell m}(\mathbf{n}), \quad (6.1)$$

where

$$a_{\ell m} = \int d\Omega Y_{\ell m}^*(\mathbf{n}) \Theta(\mathbf{n}). \quad (6.2)$$

$Y_{\ell m}(\mathbf{n})$ denote the standard spherical harmonics with $\ell = 0, 1, \dots$ and $m = -\ell, \dots, \ell$. to obtain the CMB anuglar power spectrum one combines the $a_{\ell m}$ to

$$C_{\ell}^{TT} = \frac{1}{2\ell + 1} \sum_m \langle a_{\ell m}^* a_{\ell m} \rangle. \quad (6.3)$$

6.1.2 Transfer Function and Projection Effects

The linear evolution of the curvature perturbation to the temperature fluctuations in the CMB is given by the *transfer function* $\Delta_{T\ell}(k)$. For a given curvature perturbation the multipole moments can be inferred as

$$a_{\ell m} = 4\pi(-i)^{\ell} \int \frac{d^3k}{(2\pi)^3} \Delta_{T\ell}(k) \mathcal{R}_{\mathbf{k}} Y_{\ell m}(\hat{k}). \quad (6.4)$$

The transfer function is the line-of-sight integral over physical source terms $S_T(k, \tau)$ and a geometric projection factor $P_{T\ell}(k[\tau_0 - \tau])$,

$$\Delta_{T\ell}(k) = \int_0^{\tau_0} S_T(k, \tau) P_{T\ell}(k[\tau_0 - \tau]), \quad (6.5)$$

where τ_0 is conformal time today. Generally, the transfer function has to be computed numerically. The angular power spectrum can be expressed as

$$C_{\ell}^{TT} = \frac{2}{\pi} \int dk k^2 P_{\mathcal{R}}(k) \Delta_{T\ell}^2(k). \quad (6.6)$$

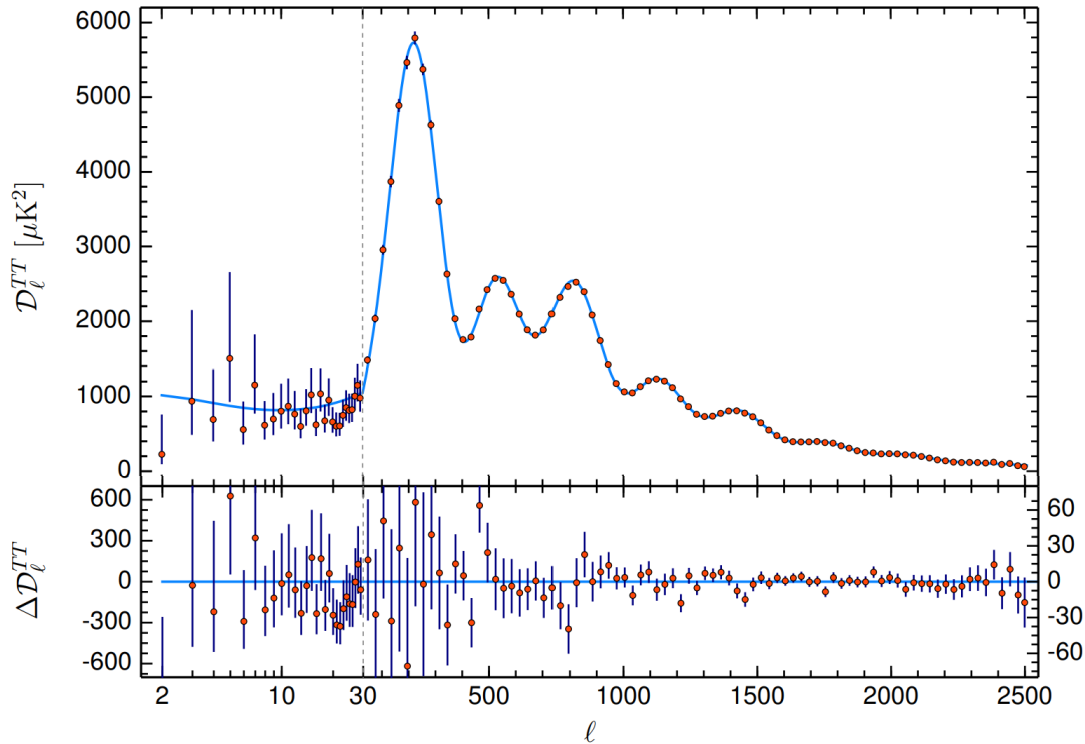


Figure 6.1: Angular power spectrum of CMB temperature fluctuations, where $\mathcal{D}_{\ell}^{TT} = \ell(\ell + 1)C_{\ell}^{TT}/2\pi$ [5].

Therefore, the CMB power spectrum is the convolution of the primordial power spectrum $P_{\mathcal{R}}$ with the anisotropies of the transfer function. To get the primordial power spectrum from the observed data, one has to invert the convolution.

6.1.3 CMB Polarisation

Before decoupling, the photons were in equilibrium with the surrounding particles via Thomson scattering. One expects the CMB to become polarised by scattering of the free electrons just before decoupling.

The E/B Decomposition

The anisotropies in the polarisation field is defined in terms of the 2×2 intensity tensor $I_{ij}(\mathbf{n})$. The components are defined relative to two orthogonal basis vectors, which are orthogonal to \mathbf{n} . The temperature anisotropies $T = \frac{1}{4}(I_{11} + I_{22})$ are invariant under a rotation in the plane of the basis vectors. Therefore, it may be expanded in terms of scalar spherical harmonics. Linear polarisation is described by the Stokes parameters

$$Q = \frac{1}{4}(I_{11} - I_{22}), \quad U = \frac{1}{2}I_{12}, \quad (6.7)$$

which transform under a rotation by an angle ψ like a spin-2 field

$$(Q \pm iU)(\hat{n}) \rightarrow e^{\mp 2i\psi}(Q \pm iU)(\hat{n}). \quad (6.8)$$

Therefore, one needs tensor spherical harmonics to describe them, or one can split them into two scalar fields

$$E(\hat{n}) = \sum_{\ell,m} a_{E,\ell m} Y_{\ell m}(\hat{n}), \quad B(\hat{n}) = \sum_{\ell,m} a_{B,\ell m} Y_{\ell m}(\hat{n}) \quad (6.9)$$

and expand them in the usual scalar spherical harmonics. These are called the E- and B-modes, respectively. While the E-mode polarisation is *curl-free* the B-modes are *divergence free*. In analogy to the temperature fluctuation one can define the angular spectra for the fluctuations of the E- and B-modes

$$C_{\ell}^{XY} = \frac{1}{2\ell + 1} \sum_m \langle a_{X,\ell m}^* a_{Y,\ell m} \rangle, \quad X, Y = T, E, B, \quad (6.10)$$

where now also cross-correlation between temperature and E-modes is possible. The other cross-correlations vanish due to symmetry reasons. Before only the curvature fluctuations \mathcal{R} acted as a source, but now also the tensor fluctuations t are relevant. The multipole moments can be concisely expressed as

$$a_{X,\ell m} = 4\pi(-i)^{\ell} \int \frac{d^3k}{(2\pi)^3} \Delta_{X\ell(k)} \{\mathcal{R}_k, t_k\} Y_{\ell m}(\hat{k}). \quad (6.11)$$

Remarkable Facts

- i) scalar (density) perturbations create only E-modes and *no* B-modes
 - ii) vector (vorticity) perturbations create mainly B-modes.
 - iii) tensor (gravitational wave) perturbations create both E- and B-modes.
- The angular power spectrum of CMB B-modes is related to the primordial tensor power spectrum $P_t(k)$ as follows

$$C_\ell^{BB} = (4\pi)^2 \int k^2 dk P_t(k) \Delta_{B\ell}^2(k). \quad (6.12)$$

Since no other early universe mechanism produces a background of tensor fluctuations that span superhorizon scales at recombination and that B-modes can only be created by tensor fluctuation implies, that the detection of B-modes would be a very strong hint that inflation occurred.

6.1.4 Coherent Phases and Superhorizon Fluctuations**The Peaks of the TT Spectrum**

The power spectrum has structure with clearly visible peaks and is not just white noise. This can only happen, when all modes of the same wavelength also have the same phase, so they can interfere coherently. Considering a Fourier mode of \mathcal{R} with wavenumber k , which is conserved outside the horizon ($k < aH$). These modes source the density fluctuation when they re-entered the horizon. The density fluctuations at the last-scattering surface lead to the fluctuations in the CMB. In order to explain the structure of the CMB power spectrum one needs a mechanism, which produces coherent initial phases for all Fourier modes. Inflation can be this mechanism, because the phases of the relevant modes were already set well before they entered the horizon, because they froze while exiting the horizon during inflation.

The Peaks of the TE Spectrum

The peaks in the TT spectrum are at $l > 200$, corresponding to angular scales $\theta < 1^\circ$. The scales have been inside the horizon when recombination happened. Therefore, it would be in principle possible to construct a mechanism which achieves the CMB peaks without using inflation. However, if one considers the TE-spectrum there is a negative peak around $100 < l < 200$, see Fig. 6.2, which corresponds to scales outside the horizon at the time of recombination. Hence, there is no possibility that this could have been produced after $\tau = 0$ and one needs something like inflation which extends conformal time into the negative regime. To get the phases aligned one needs a primordial mechanism, when they were still in causal contact.

6.2 21cm Intensity Mapping

Another way to probe the inflationary mechanism is to use the distribution of neutral hydrogen (HI) in the universe. With the knowledge of where the hydrogen is located one can infer the distribution of all matter and get the matter power spectrum. The section closely follows Ref.[27, 28]. The mapping of neutral hydrogen

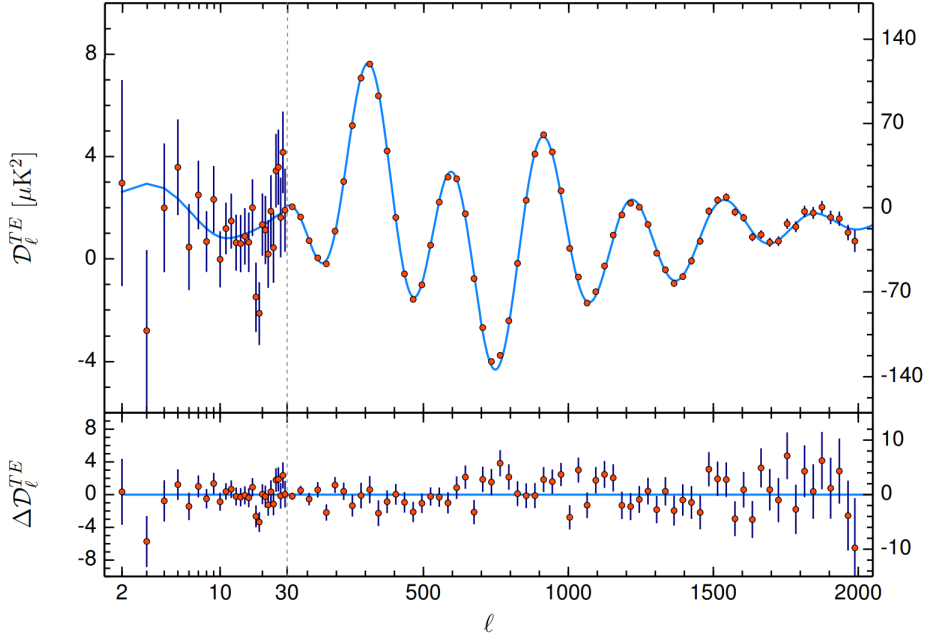


Figure 6.2: Power spectrum of the cross-correlation between temperature and E-mode polarisation anisotropies, where $\mathcal{D}_\ell^{TE} = \ell(\ell + 1)C_\ell^{TE}/2\pi$ [5].

has great potential, as allows to find a three dimensional map of the matter density fluctuations as compared to the 2D slice of the CMB. The significance of this aspect is illustrated in Fig. 6.3. Neutral hydrogen can be observed through the 21cm line. It is generated by the forbidden spin-flip transition between the singlet and the triplet state. Its energy corresponds to a wavelength of $\lambda = 21.106$ cm, hence the name. This transition is highly forbidden with transition rate of $2.9 \cdot 10^{-15} \text{ s}^{-1}$. However, due to the sheer abundance of hydrogen in the universe the signal can still be observed.

The line can be observed in emission or absorption against the CMB, depending on the spin temperature of the gas. During the evolution of the universe, the difference between these temperature varies quite a lot and the different phases are schematically illustrated in Fig. 6.4. Most significantly during the cosmic dawn, where the first galaxies and stars started to form as well as during the process of reionization. It describes the period in the evolution the universe, where the newly formed stars ionized the hydrogen around them. This is not an instantaneous process, which means, that the fraction of neutral hydrogen in the universe is redshift dependent.

The spin temperature of hydrogen T_S is defined by

$$\frac{n_1}{n_0} = 3e^{-T_*/T_S}, \quad (6.13)$$

where n_0 and n_1 describe the abundance of the singlet and triplet state, respectively. $T_* = 5.9 \mu\text{eV}$ is the energy difference between the hyperfine levels. The spin temperature is coupled to the gas temperature through the Wouthuysen-Field effect, where Lyman- α photons excite the hydrogen, which might then decay into the other spin state. The gas in the intergalactic medium (IGM) is heated by X-ray photons [30]. The local brightness temperature is given by

$$T_{21}^{loc} = (T_S - T_{\text{cmb}})(1 - e^{-\tau}), \quad (6.14)$$

where τ is the optical depth, depending on the local neutral hydrogen density [31]. Now, one can either use approximate calculations, to get to the matter power spectrum or one can use a numerical approach to obtain the full, so called light-cone. A lightcone is a 3D object, with two spatial directions and the redshift, where one has the 21cm temperature. It is obtained by simulating the evolution of densities using perturbation theory. This work is mainly focused on the power spectrum, as it has a lot less parameters, but in Ch. 9 the full light-cones are used, where the production is discussed in more detail.

Continuing in the power spectrum approach, adjusting for the redshift and using the optical thinness of the IGM ($\tau \ll 1$), as well as more simplifications given in Ref.[28], yields

$$\Delta T_{21} \approx 189 \left[\frac{H_0(1+z)^2}{H(z)} \right] \Omega_{\text{HI}}(z) h \text{ mK}, \quad (6.15)$$

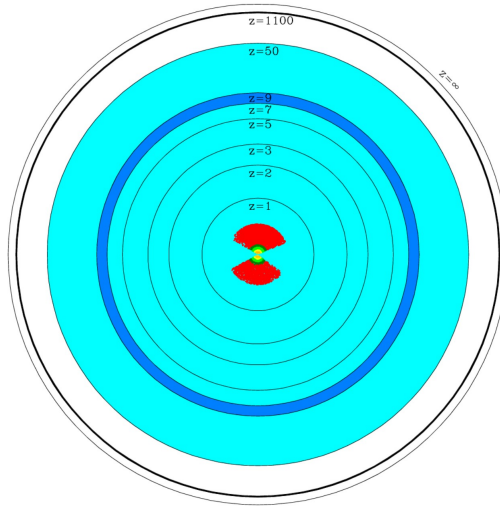


Figure 6.3: Volume of the Universe, that can be mapped with 21cm tomography (cyan) [28].

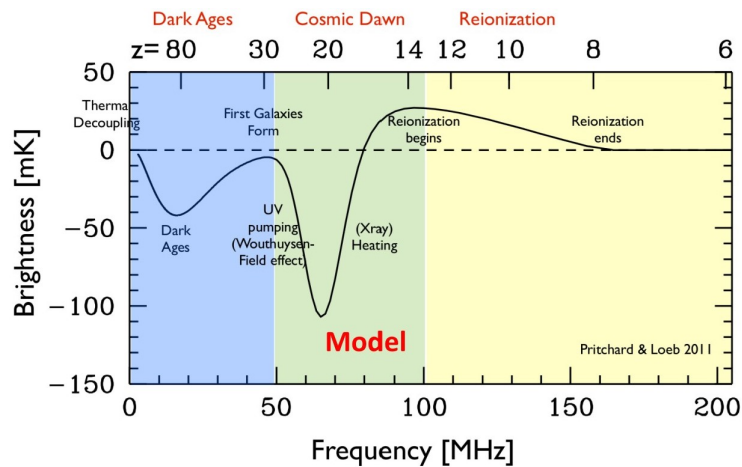


Figure 6.4: Illustration of the 21cm brightness temperature as a function of redshift [29].

with the reduced Hubble-function h and the mass density of neutral hydrogen

$$\Omega_{\text{HI}} = \frac{\rho_{\text{HI}}}{\rho_c} = \Omega_b(1 - Y_P) \left(\frac{H_0}{H(z)} \right)^2 (1 + z)^3 x_{\text{HI}}. \quad (6.16)$$

In this thesis, the considered redshift window is $z = 8 \dots 10$, there the astrophysics of reionization is manageable in terms of simple modelling. Yet, reionization has not ended and the signal is still observable. This allows for the following empirical fitting formula for the z -dependence of neutral hydrogen fraction [32]

$$x_{\text{HI}}(z) = \frac{1}{2} \left[1 + \frac{2}{\pi} \tan^{-1}(\delta_1(z - \delta_2)) \right]. \quad (6.17)$$

The two parameters δ_1 and δ_2 are used to account for some variation in the astrophysics leading to a slightly changed neutral hydrogen fraction. In the analysis they are going to be nuisance parameters.

As mentioned in the introduction to this section, the 21cm power spectrum P_{21} is related to the matter power spectrum P_δ , which in turn can be calculated from the primordial power spectrum $P_{\mathcal{R}}$ as

$$P_\delta(k, \tau) = \frac{4}{25} \left(\frac{k}{aH} \right)^4 T_\delta^2(k, \tau) P_{\mathcal{R}}, \quad (6.18)$$

where $\delta = \delta\rho/\bar{\rho}$ and T_δ is the matter transfer function. In this thesis the transfer function will be calculated numerically with CLASS, discussed in Sec. 7.6.

The 21cm power spectrum P_{21} is defined through the relation

$$\langle \Delta T_{21}(\mathbf{k}) \Delta T_{21}(\mathbf{k}') \rangle \equiv P_{21}(\mathbf{k}, z) (2\pi)^3 \delta(\mathbf{k} - \mathbf{k}'). \quad (6.19)$$

It is related to the matter power spectrum P_δ in the flat-sky approximation [33] by

$$P_{21}(k, \mu, z) = f_{\text{obs}} \times b_{21}^2(z) \times P_\delta(\hat{\mu}, z), \quad (6.20)$$

with the Fourier mode \mathbf{k} and the line-of-sight \mathbf{r} describing

$$k = |\mathbf{k}| \quad \text{and} \quad \mu = \frac{\mathbf{k} \cdot \mathbf{r}}{kr}. \quad (6.21)$$

The hatted quantities are the ones, derived from the true cosmology and the ones without a hat are derived from the fiducial cosmological model, which is probed.

The bias term $b_{21}(z)$ is given by

$$b_{21} = \Delta T_{21}(z) b_{\text{HI}}(z), \quad (6.22)$$

where ΔT_{21} is given in Eq. (6.15) and the bias b_{HI} will be treated as nuisance parameter. Now, the term describing observational effects f_{obs} of Eq. (6.20) consists of the following parts

$$f_{\text{obs}} = f_{\text{AP}}(z) \times f_{\text{res}}(k, \mu, z) \times f_{\text{RSD}}(\hat{k}, \hat{\mu}, z) \quad (6.23)$$

and will be discussed one by one. Starting with the Alcock-Paczynski effect, which accounts for the relative change in the power spectrum between true cosmology and the assumed one

$$f_{\text{AP}}(z) = \frac{D_A^2 \hat{H}}{\hat{D}_A^2 H}, \quad (6.24)$$

with the Hubble parameter H and the angular diameter distance D_A .

Next, f_{res} describes the final resolution of the instruments, measuring the signal. Its explicit formula is

$$f_{\text{res}}(k, \mu, z) = \exp \left[-k^2 (\mu^2 (\sigma_{\parallel}^2 - \sigma_{\perp}^2) + \sigma_{\perp}^2) \right], \quad (6.25)$$

where σ_{\parallel} and σ_{\perp} are the Gaussian errors of the coordinates parallel and perpendicular to the line-of-sight at redshift z . They can be calculated as

$$\sigma_{\parallel} = \frac{c}{H} (1+z)^2 \frac{\sigma_{\nu}}{\nu_0} \quad \text{and} \quad \sigma_{\perp} = (1+z) D_A \sigma_{\theta}, \quad (6.26)$$

$$\text{with} \quad \sigma_{\theta} = \frac{1}{\sqrt{8 \ln 2}} \frac{\lambda_0}{D_{\text{base}}} (1+z) \quad \text{and} \quad \sigma_{\nu} = \frac{\delta_{\nu}}{\sqrt{8 \ln 2}}. \quad (6.27)$$

The experimental values of the parameters and their meaning are given in Tab.7.2 in Sec. 7.5.1.

The next term corrects an apparent anisotropy in the power spectrum arising from the cosmological redshift as well as fingers-of-God effect [34]

$$f_{\text{RSD}}(\hat{k}, \hat{\mu}, z) = \left(1 + \beta(\hat{k}, z) \hat{\mu}^2 \right)^2 e^{-\hat{k}^2 \hat{\mu}^2 \sigma_{\text{NL}}^2}$$

$$\text{with} \quad \beta(\hat{k}, z) = -\frac{1+z}{2b_{21}(z)} \frac{d \log P_{\delta}(\hat{k}, z)}{dz}. \quad (6.28)$$

CHAPTER 7

Statistics and Tools

7.1 Bayesian Inference

Physics works by predicting mathematically consistent theories and comparing them to observations, discarding those, which are not compatible with data. Deriving statistical statements about the validity of a theory is called *inference*. However, every observation has some statistical error as the amount of data is always finite and due to the naturalness of experiments. The experiment itself is part of nature and can therefore also only be described by the current best model. Hence, one can only differentiate theories which are statistically not compatible with the observation. This brief introduction to Bayesian-inference is based on Ref.[35].

The likelihood $\mathcal{L}(\{y_i\}|\theta_\mu)$ describes how probable it is to observe the data points $y = \{y_i\}$ given a set of model parameters $\theta_\mu = (\theta_1, \dots, \theta_n)$. Then, the set of parameters, which maximise \mathcal{L} are an estimator of the true parameters, as they are most likely produce the observed data. This principle is called *maximum likelihood*.

Unfortunately, that is not what one is particularly interested in. One would actually like to know the distribution of the model parameters given a set of data, $p(\theta_\mu|\{y_i\})$. To interchange the random variable and the condition, one has to use Bayes-theorem:

$$p(\theta_\mu|\{y_i\}) = \frac{\mathcal{L}(\{y_i\}|\theta_\mu)}{p(\{y_i\})}p(\theta_\mu) \quad (7.1)$$

The *posterior* distribution $p(\theta_\mu|\{y_i\})$ is given by the product of the *prior* distribution $p(\theta_\mu)$, i.e. how likely is the model before carrying out the experiment, and the likelihood, normalised by the evidence $p(\{y_i\})$. The evidence is the probability of obtaining the data,

$$p(\{y_i\}) = \int d^n\theta \mathcal{L}(\{y_i\}|\theta_\mu)p(\theta_\mu). \quad (7.2)$$

Often, one is interested in the posterior of one explicit parameter, independent of the rest of the set. This can be achieved by a process called *marginalisation*, which is essentially integrating out the other parameters, e.g.

$$p(\theta_1|\{y_i\}) = \int p(\theta_\mu|\{y_i\})d\theta_2 \dots d\theta_n. \quad (7.3)$$

In principle it is possible to numerically compute the posterior distribution for a given model and a given data set. However, that can be very challenging due to the possibly highly-dimensional parameter space. To manage this problem, one uses some kind of *Markov chain Monte Carlo* (MCMC) to generate samples θ_μ , which are distributed according to $p(\theta_\mu|\{y_i\})$.

7.2 Metropolis-Hastings algorithm

The basic idea of MCMC is to draw values of θ (discarding the subscript μ for readability) from a distribution and then correcting those draws to get a better approximation. Each draw θ^t only ever depends on the previous draw θ^{t-1} , which is the Markov property. The distribution is improved at each step, i.e. it is converging to the target distribution $p(\theta|\{y_i\})$.

A popular algorithm to do that is the Metropolis-Hastings algorithm [36]. It proceeds as follows:

1. Draw a starting point θ^0 , with $p(\theta^0|y) > 0$ from a starting distribution $p_0(\theta)$.
2. For $t = 1, 2, \dots$:
 - (a) Sample a proposal θ^* from a *jumping distribution* $J_t(\theta^*|\theta^{t-1})$ at time t .
 - (b) Calculate the ratio of densities,

$$r = \frac{p(\theta^*|y)/J_t(\theta^*|\theta^{t-1})}{p(\theta^{t-1}|y)/J_t(\theta^{t-1}|\theta^*)}. \quad (7.4)$$

- (c) Set

$$\theta^t = \begin{cases} \theta^* & \text{with probability } \min(r, 1) \\ \theta^{t-1} & \text{otherwise} \end{cases} \quad (7.5)$$

Step 2 is repeated until the chain has reached the desired length. To summarise the algorithm one can state the acceptance/rejection rule by: (a) if the jump increases the posterior density, set $\theta^t = \theta^*$; (b) if the jump decreases the posterior density, set $\theta^t = \theta^*$ with probability equal to the density ratio r , and set $\theta^t = \theta^{t-1}$ otherwise.

By construction the posterior is not known, but one can use Bayes-theorem to rewrite

$$\frac{p(\theta^*|y)}{p(\theta^{t-1}|y)} = \frac{\mathcal{L}(y|\theta^*)p(\theta^*)}{\mathcal{L}(y|\theta^{t-1})p(\theta^{t-1})}, \quad (7.6)$$

where likelihood and prior are known and the evidence cancels out.

Since one starts from a random point in the parameter space it takes some time for the chain to equilibrate. Therefore, one discards the first few iterations, which is called the *burn-in*, before analysing the rest of the chain.

There are arbitrary many possible jumping distributions, however there are good and bad ones. A sensible jumping distribution, or sometimes called proposal distribution, has the following properties:

- It is easy to sample any θ from $J(\theta^*|\theta)$.
- It is easy to compute the ratio r .

- Each jump goes a reasonable distance in the parameter space, otherwise the random walk is too slow.
- The proposals are not rejected too frequently, otherwise the random walk wastes too much time standing still.

It is a very challenging task to check whether one has reached convergence or not. One method to do so is the $(R - 1)$ criterium or Gelman-Rubin statistic [37]. For this one needs to run multiple chains and it compares the variance within each chain with the variance between multiple chains, explicitly

$$\bar{x}_j = \frac{1}{L} \sum_{t=1}^L x_t^{(j)} \quad (\text{chain mean}) \quad (7.7)$$

$$\bar{x} = \frac{1}{J} \sum_{j=1}^J \bar{x}_j \quad (\text{grand mean}) \quad (7.8)$$

$$s_j^2 = \frac{1}{L-1} \sum_{t=1}^L (x_t^{(j)} - \bar{x}_j)^2 \quad (\text{within chain variance}) \quad (7.9)$$

$$B = \frac{L}{J-1} \sum_{j=1}^J (\bar{x}_j - \bar{x})^2 \quad (\text{between chain variance}) \quad (7.10)$$

$$W = \frac{1}{J} \sum_{j=1}^J s_j^2 \quad (7.11)$$

and the Gelman-Rubin statistic is then

$$R = \frac{\frac{L-1}{L}W + \frac{1}{L}B}{W}. \quad (7.12)$$

If a set of chains is converged, the between and within chain variance become equal, and for infinite chain length $R \rightarrow 1$. Therefore, if $R - 1 \ll 1$ one considers the chains converged. In this work, $R - 1 < 0.05$ is required.

7.3 Toy Example

In order to illustrate how MCMC works this sections presents a very simple toy example. The likelihood is given by

$$\mathcal{L}(y|\theta_1, \theta_2) \propto \exp\left(-\frac{(y - \sqrt{\theta_1^2 + \theta_2^2})^2}{2}\right), \quad (7.13)$$

where θ_i are the parameter of the model and y the data. The normalisation is irrelevant as one only needs ratios. This likelihood is just a two dimensional ring with radius y and Gaussian thickness.

As a proposal distribution a simple Gaussian with $\mu = 0$ and two different values of $\sigma_{\text{prop}} = \{0.3, 3\}$ are chosen. Due to the symmetry of this distribution, it cancels in the calculation of r in Eq. (7.4). To make it even simpler, the prior is for both

parameters flat. Then, only the likelihood ratio enters the calculation if the new point in the chain is accepted. For the simulation $y = 4$ is set and the starting point is purposely chosen far away to show the need for a burn-in phase.

In Fig. 7.1a the evolution of a chain with small jumps is shown. The small jumps come from the small width of the proposal distribution. One can clearly see the starting point and the walk towards the minimum of the likelihood, i.e. the ring. Once the chain reaches the ring, it wanders around in circles, shown by the distinct colour patches. This implies, that the chain has still not properly converged and one needs a lot more steps. In Fig. 7.1c the marginalised posteriors for the same set-up as in Fig. 7.1a are plotted, but now for up to 100 times more steps. After 10^4 steps the chain is definitely not converged, as it still has a strong asymmetry, but after 10^6 steps, the chain had enough time to explore the whole parameter space and thermalise. The blue curve is the correct posterior for this toy example. Lastly, in Fig. 7.1b the evolution of a chain is shown, however, with a wider proposal distribution. This chain thermalises a lot quicker, as now it does not have to wander around in the valley to reach the other side of the ring, it can just jump over the maximum in the middle. The acceptance rate of the configuration is smaller, because a lot of the large jumps are rejected, but in the end, the equilibrium is reached faster.

7.4 MontePython

The tool to perform the MCMC analysis in this work is MontePython [38]. It is a Python package designed to do parameter inference in cosmology. The code is able to explore the parameter space with multiple techniques, e.g. Nested Sampling, ensemble MCMC and Metropolis-Hastings. In this work an improved version of the Metropolis-Hastings algorithm, described in Sec. 7.2, is used. The jumping distribution is just the old point plus a random jump drawn from a Gaussian proposal distribution of the form

$$p = \mathcal{N} \exp \left\{ -\frac{1}{2c} \Delta \mathbf{p}^T \mathbf{C}^{-1} \Delta \mathbf{p} \right\}, \quad (7.14)$$

where $\Delta \mathbf{p}$ are the parameter jumps and \mathbf{C} is a matrix with the information about the parameter correlation and the standard deviations relative to each other. The actual covariance matrix is given by $c\mathbf{C}$, with the jumping parameter c .

For an optimal proposal distribution one should get an acceptance rate of around 0.25, which can be achieved for Gaussian posterior distributions if the covariance matrix is approximately the one of the posterior and the jumping parameter is 2.4^2 [39]. To optimize the proposal function the code has a feature, which allows to automatically update the covariance matrix. This mechanism considers the worst parameter, i.e. $\max(R - 1)$ and only updates if it is below 3 and above 0.4. The upper bound is implemented to not make the covariance matrix worse due to a statistical coincidence. The lower one is chosen to not make an unnecessary update which might improve the convergence speed marginally but the steps prior to the update have to be thrown away as they are no longer Markovian. At some point it helps more with convergence to accumulate more steps and not have a slightly better covariance matrix. The points are no longer Markovian, because they are

used to create a new proposal function and hence, the next point depends on more than the previous step.

MontePython is interfaced with CLASS (see. Sec. 7.6), which allows to calculate the desired observables from the set of parameters proposed by MontePython. The MontePython package includes several likelihoods such as the Planck likelihoods [5] and multiple CMB mock-likelihoods [40], which are discussed in Sec. 7.5.1.

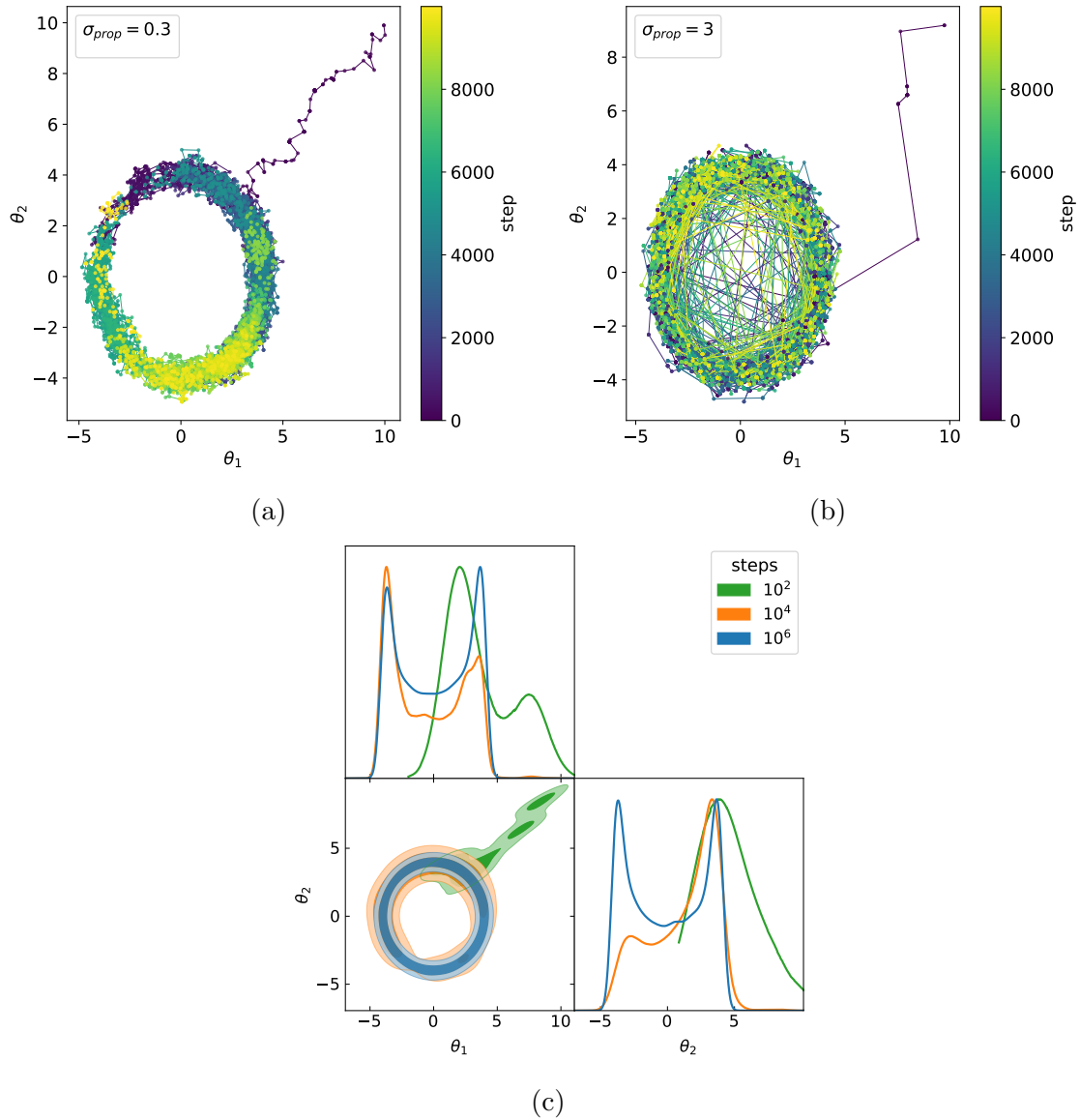


Figure 7.1: Markov chain Monte Carlo simulation for the likelihood in Eq. (7.13). In (a) and (b) each step in the evolution of the chain is shown, where the difference in the two plots is the width of the proposal distribution. Plot (c) shows the marginalised posterior at different times for the same configuration as in (a) but for much more steps.

7.5 Likelihoods

The likelihood is the central object in any MCMC analysis. Since, there is only data from the Planck-satellite one needs to construct a so called *mock-likelihood* for the desired forecasts. In this work there are two different classes of mock-likelihoods considered. First, for future CMB-experiments and second for the square kilometer array (SKA).

7.5.1 CMB

The likelihoods for future CMB-experiments, explicitly LiteBIRD and CMB-S4, are already implemented in MontePython and the discussion will follow Ref.[40]. The data a CMB-experiment will collect are best expressed in spherical harmonics and receive a contribution from signal and noise,

$$a_{\ell m}^X = s_{\ell m}^X + n_{\ell m}^X, \quad (7.15)$$

where $X = T, E, B$ as discussed in the CMB section 6.1. The noise can be modelled as a Gaussian beam with an additional spatially uniform Gaussian white noise.

$$N_{\ell}^{XX'} = \langle n_{\ell m}^{X*} n_{\ell m}^{X'} \rangle = \delta_{XX'} \theta_{\text{fwhm}}^2 \sigma_X^2 \exp \left[\ell(\ell + 1) \frac{\theta_{\text{fwhm}}^2}{8 \ln 2} \right], \quad (7.16)$$

where θ_{fwhm} and σ_X depend on the experiment and are listed in Tab. 7.1. Off diagonal noise terms are expected to vanish, as it is expected to be uncorrelated. The primary CMB photons interact with various different astrophysical objects and the signal has to be adjusted. This process is called foreground cleaning. After which, the signal and noise are uncorrelated and the total power spectrum is given by

$$\langle a_{\ell m}^{X*} a_{\ell' m'}^{X'} \rangle = \left(C_{\ell}^{XX'} + N_{\ell}^{XX'} \right) \delta_{\ell\ell'} \delta_{mm'}. \quad (7.17)$$

Given some fiducial model the power spectra can be calculated with CLASS. The mock data is obtained by following a simple algorithm:

- (i) Generate Gaussian distributed random numbers $G_{\ell m}^{(i)}$ with unit variance.
- (ii) Define the T , E and B multipole moments as

$$a_{\ell m}^T = \sqrt{\bar{C}_{\ell}^{TT}} G_{\ell m}^{(1)}, \quad (7.18)$$

$$a_{\ell m}^E = \frac{\bar{C}_{\ell}^{TE}}{\bar{C}_{\ell}^{TT}} \sqrt{\bar{C}_{\ell}^{TT}} G_{\ell m}^{(1)} + \sqrt{\bar{C}_{\ell}^{EE} - \frac{(\bar{C}_{\ell}^{TE})^2}{\bar{C}_{\ell}^{TT}}} G_{\ell m}^{(2)}, \quad (7.19)$$

$$a_{\ell m}^B = \frac{\bar{C}_{\ell}^{TB}}{\bar{C}_{\ell}^{TT}} \sqrt{\bar{C}_{\ell}^{TT}} G_{\ell m}^{(1)} + \sqrt{\bar{C}_{\ell}^{BB} - \frac{(\bar{C}_{\ell}^{TB})^2}{\bar{C}_{\ell}^{TT}}} G_{\ell m}^{(3)}, \quad (7.20)$$

where $\bar{C}_{\ell}^{XX'} \equiv C_{\ell}^{XX'} + N_{\ell}^{XX'}$ is the fiducial signal plus the noise.

With the mock data the likelihood for the MCMC analysis is simply given by

$$\mathcal{L}(\mathbf{a}|\Theta) \propto \exp \left(-\frac{1}{2} \mathbf{a}^{\dagger} \bar{C}(\Theta)^{-1} \mathbf{a} \right), \quad (7.21)$$

	Channel [GHZ]	FWHM [arcmin]	ΔT [$\mu\text{K arcmin}$]	f_{sky}	ℓ range
LiteBIRD	140	31	4.1	0.7	2 - 1350
CMB-S4	150	3.0	1.0	0.4	30 - 3000

Table 7.1: Experimental specifications for the CMB experiments used in this work.

where $\mathbf{a} = (a_{\ell m}^T, a_{\ell m}^E, a_{\ell m}^B)$ is the data vector and Θ describes the theoretical model. Usually, one uses the χ^2 functional to describe the likelihood. It is given by

$$\chi_{\text{eff}}^2 = \sum_{\ell} (2\ell + 1) f_{\text{sky}} \left(\frac{D}{|\bar{C}|} + \ln \frac{|\bar{C}|}{|\hat{C}|} - 3 \right), \quad (7.22)$$

where $|\bar{C}|$ and $|\hat{C}|$ are the determinants of the theoretical and observed data covariances, respectively. D is an elaborate combination of the spectra and given in [40]. The multiplicative factor $f_{\text{sky}} < 1$ is introduced because no experiment can cover the full sky. The normalisation of χ_{eff}^2 is chosen such, that it becomes zero when theoretical and observed spectra are equal.

The description of the mock data and likelihood requires input from experimental specifications of the telescopes. This work focuses on the future CMB experiments LiteBIRD [10, 11] and CMB-S4 [15, 14]. Their specifications are listed in Tab. 7.1 [41].

LiteBIRD

The LiteBIRD satellite is designed to measure the B-mode polarisation of the CMB. On large angular scales the B-modes carry the imprint of primordial gravitational waves, whose measurement would provide a powerful probe of the epoch of inflation. The large sky coverage of this satellite makes it a perfect candidate to achieve this goal.

CMB-S4

CMB-S4 is a ground based CMB-telescope with various science goals. One of them is to detect B-mode polarisation. Contrary to LiteBIRD it does not have a huge sky coverage, but the far better resolution due to it being on the ground and the technical advantages that carries.

These two experiments have very similar goals but are quite different in their approaches. This makes them perfect for combining their data. LiteBIRD is superior on large angular scales and CMB-S4 on small ones.

7.5.2 Square Kilometer Array

In order to be able to forecast how the future 21cm experiment *Square Kilometer Array* (SKA) will constrain inflationary models one needs a mock-likelihood. The discussion of its construction follows closely Ref.[32].

SKA is a radio telescope consisting of thousands of small antennas spread out over

an area of approximately one square kilometer. The experiment is based in Australia and in South Africa, where each site has different experimental specifications. This thesis uses for the analysis the specifications for SKA-low, listed in Tab. 7.2, which is the one based in South Africa [18].

The spectra of the primordial tensor and scalar perturbations P_t and $P_{\mathcal{R}}$ are evolved with CLASS yielding the matter power spectrum $P_{\delta}(k)$. The relation between the 21cm power spectrum and the matter power spectrum is given by Eq. (6.20).

The χ^2 defining the likelihood is given by

$$\chi^2 = \sum_{\text{bins } n} \int_{k_{\min}}^{k_{\max}} k^2 dk \int_{-1}^1 d\mu \frac{V_r(\bar{z}_n)}{2(2\pi)^2} \left[\frac{(\Delta P_{21}(k, \mu, \bar{z}_n))^2}{(P_{21}(k, \mu, \bar{z}_n) + P_N)^2 + \sigma_{\text{th}}^2(k, \mu, \bar{z}_n)} \right], \quad (7.23)$$

where the redshift range is divided into $n = 21$ bins and each bin has its mean redshift \bar{z}_n and an approximated volume of

$$V_r(\bar{z}) = 4\pi f_{\text{sky}} \int_{\Delta r(\bar{z})} r^2 dr = \frac{4\pi}{3} f_{\text{sky}} \left[r^3 \left(\bar{z} + \frac{\Delta z}{2} \right) - r^3 \left(\bar{z} - \frac{\Delta z}{2} \right) \right]. \quad (7.24)$$

ΔP_{21} is the difference between the fiducial and the sampled power spectra, see Sec. 6.2. Now, there are two unspecified terms in Eq. 7.23 left, namely, the noise of the power spectrum P_N and the theory uncertainty σ_{th} . The latter depends on the correlation length ($\Delta k, \Delta \mu, \Delta z$) is discussed in detail in Ref.[32].

The formula for the noise reads [42]

$$P_N(z) = \frac{4\pi T_{\text{sys}}^2 f_{\text{sky}} \lambda^2 y D_A^2}{A \Omega f_{\text{cover}} t_{\text{obs}}}, \quad (7.25)$$

where all the experimental specifications are given in Tab. 7.2. The system temperature is a combination of sky temperature and the receiver temperature

$$T_{\text{sys}} = T_{\text{sky}} + T_{\text{rx}}$$

with $T_{\text{sky}} = 25 \text{ K} \left(\frac{408 \text{ MHz}}{\nu} \right)^{2.75}$ and $T_{\text{rx}} = 0.1 T_{\text{sky}} + 40 \text{ K}$. (7.26)

Finally, y is defined as

$$y = \frac{18.5 \text{ Mpc}}{1 \text{ MHz}} \left(\frac{1+z}{10} \right)^{1/2}. \quad (7.27)$$

7.6 CLASS

The Boltzmann-Code CLASS (The Cosmic Linear Anisotropy Solving System [43]) is used in two ways in this analysis. First, the possibility to calculate the primordial power spectrum from an arbitrary inflationary potential is implemented. The potential only has to fulfil a few conditions: It has to be positive, its slope negative and one can write down an analytic expression for it and its first two derivatives. To calculate the primordial power spectrum CLASS uses the equations from Sec. 5.4.

Parameter	Value/Formula	Explanation
λ_0	21.11 cm	wavelength
ν_0	1420.405752 MHz	frequency
D_{base}	1 km	maximal baseline
D	40 m	antenna diameter
t_{obs}	10000 hrs	total observation time
N_{dish}	224	number of antennas
f_{sky}	0.58	sky coverage
z	8 ... 10	considered redshift range
k	0.01 ... 0.2 Mpc^{-1}	wavenumber
Ω	$1.2\lambda/D)^2$	field of view
A	$N_{\text{dish}}\pi(D/2)^2$	total area
f_{cover}	$N_{\text{dish}}(D/D_{\text{base}})^2$	covering fraction

Table 7.2: Experimental specifications of SKA1-LOW [18]

From the primordial power spectrum CLASS is able to time evolve the perturbations up to a desired redshift using the transfer functions from Sec. 6.1 and 6.2. It can then calculate the angular power spectrum a CMB-experiment would measure at $z = 1100$ as well as the matter power spectrum at e.g. $z = 9$, as shown in Fig. 7.2.

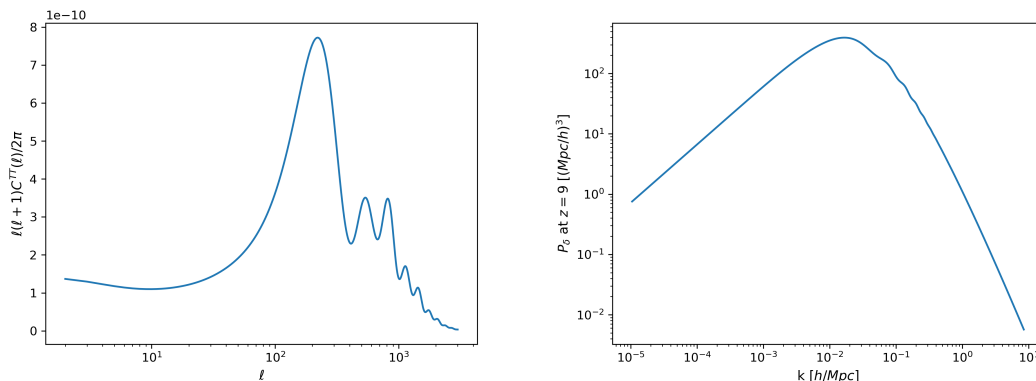


Figure 7.2: Example angular and matter power spectra calculated with CLASS using the Planck best-fit values.

CHAPTER 8

Results and Forecasts

Finally, all the theory and methods are introduced and it is possible to discuss the analysis. To perform the analysis one needs to decide on which parameters describe the model. Here, the set consisting of

$$\{\omega_b, \omega_{\text{cdm}}, h, \tau_{\text{reio}}, M, c, N_\star\} \quad (8.1)$$

is chosen. Where the first four parameters describe the cosmology and the last three describe the inflationary scenario:

- Baryon density today $\omega_b = \Omega_b h^2$
- Cold dark matter density today $\omega_{\text{cdm}} = \Omega_{\text{cdm}} h^2$
- Reduced Hubble constant today $h = H_0/100$
- Optical thickness at reionization τ_{reio}
- First Starobinsky parameter M
- Second Starobinsky parameter c
- e -folds of inflation after the reference mode exited the horizon N_\star

Even though the mechanism for inflation alters standard GR the evolution of the Universe is considered in a FLRW-cosmology, because the higher curvature terms become negligible after inflation.

Given this parameter set CLASS calculates the desired spectra, MontePython calculates the likelihood for this set and then proposes a new set according to the rules of the Metropolis-Hastings algorithm, described in Section 7.2. All parameters have a flat prior except N_\star , for which a Gaussian one with $\mu = 55$ and $\sigma = 5$ was chosen as the chains converge considerably faster and the difference in the results is marginal. These values are motivated by the flatness and horizon problems. The prior for c disallows negative values, as discussed in Ch. 5. The reference scale is set to $k_\star = 0.05 \text{ Mpc}^{-1}$. To perform the MCMC analysis a total of 8 chains with up to 5.5 million steps combined are used. As convergence criterium $R - 1 < 0.05$ is chosen. All plots are made with the plotting tool GetDist [44].

This section shows the results of this procedure for the various likelihoods considered.

It starts with the data from the Planck collaboration and then forecasts for the next generation CMB experiments and the SKA are discussed.

8.1 Planck Results

The Planck data is used to reproduce the results in Ref.[5] and verify the validity of the framework used in the forecasts. The marginalised posterior distributions of the parameter set (8.1) from this analysis are shown in Fig. 8.1. The best-fit values as well as the mean and its error are given in Tab. 8.1. All of the obtained values are compatible with the Planck results. The best-fit values for the two most interesting parameters for this work are

$$\frac{M}{M_P} = 1.103 \cdot 10^{-5} \quad \text{and} \quad c = 4.135 \cdot 10^{-5}, \quad (8.2)$$

which are also consistent with previous work on the extension of the Starobinsky model [45, 46, 47]. The parameter c in front of the R^3 term, describing the extension of the model, is clearly compatible with 0. This is absolutely necessary, as the pure Starobinsky model is the best-fit model of the Planck data. Negative values of c are excluded, as this leads to an unphysical inflationary potential. It is worth noting, that the data shows a slight bias towards non-zero values of c and therefore, towards an extension of the Starobinsky model.

8.2 CMB Forecasts

The next step is to show what the next generation CMB experiments can do to improve the constraints on the inflationary parameters. To do so, the mock likelihoods, discussed in Sec. 7.5.1, are used. For the fiducial, or assumed true, value the best-fit parameter values from the MCMC analysis of the Planck data is used.

The forecasts are done for the two experiments, LiteBIRD and CMB-S4 alone and their consistent combination. For the combination one uses LiteBIRD for $\ell = 2 \dots 50$ and CMB-S4 for $\ell = 51 \dots 3000$ to profit from the higher sky coverage of LiteBIRD as well as the higher angular resolution of CMB-S4. It is important to note that one cannot just combine their full ℓ ranges as they both measure the same phenomenon.

The marginalised distributions are shown in Fig. 8.2 and the values in Tab. 8.1. Even though LiteBIRD and CMB-S4 are both CMB-experiments their different focus can be seen in the different constraints they produce. The superior polarisation sensitivity of LiteBIRD manifests itself in the parameter τ_{reio} . There, the constraints are a lot stronger compared to the ones from CMB-S4. However, with any other parameter CMB-S4 has the upper hand. For example, ω_b is extracted from alternating peak heights of the acoustic peaks. Naturally, CMB-S4 yields a better measurement due to the large number of multipoles it can probe. Fortunately, one can consistently combine these to and get the best of both worlds, as is shown by the yellow contours in Fig. 8.2. This combination would allow a measurement of the assumed

true value of $M = 1.119 \cdot 10^{-5} M_P$ and $c = 4.135 \cdot 10^{-5}$ to

$$\frac{M}{M_P} = (1.064 \dots 1.147) \cdot 10^{-5} \quad (95\%CL) \quad (8.3)$$

$$c = (1.015 \dots 8.3) \cdot 10^{-5} \quad (95\%CL) , \quad (8.4)$$

and therefore show at a 2σ level, that an extension of the Starobinsky model of inflation is necessary.

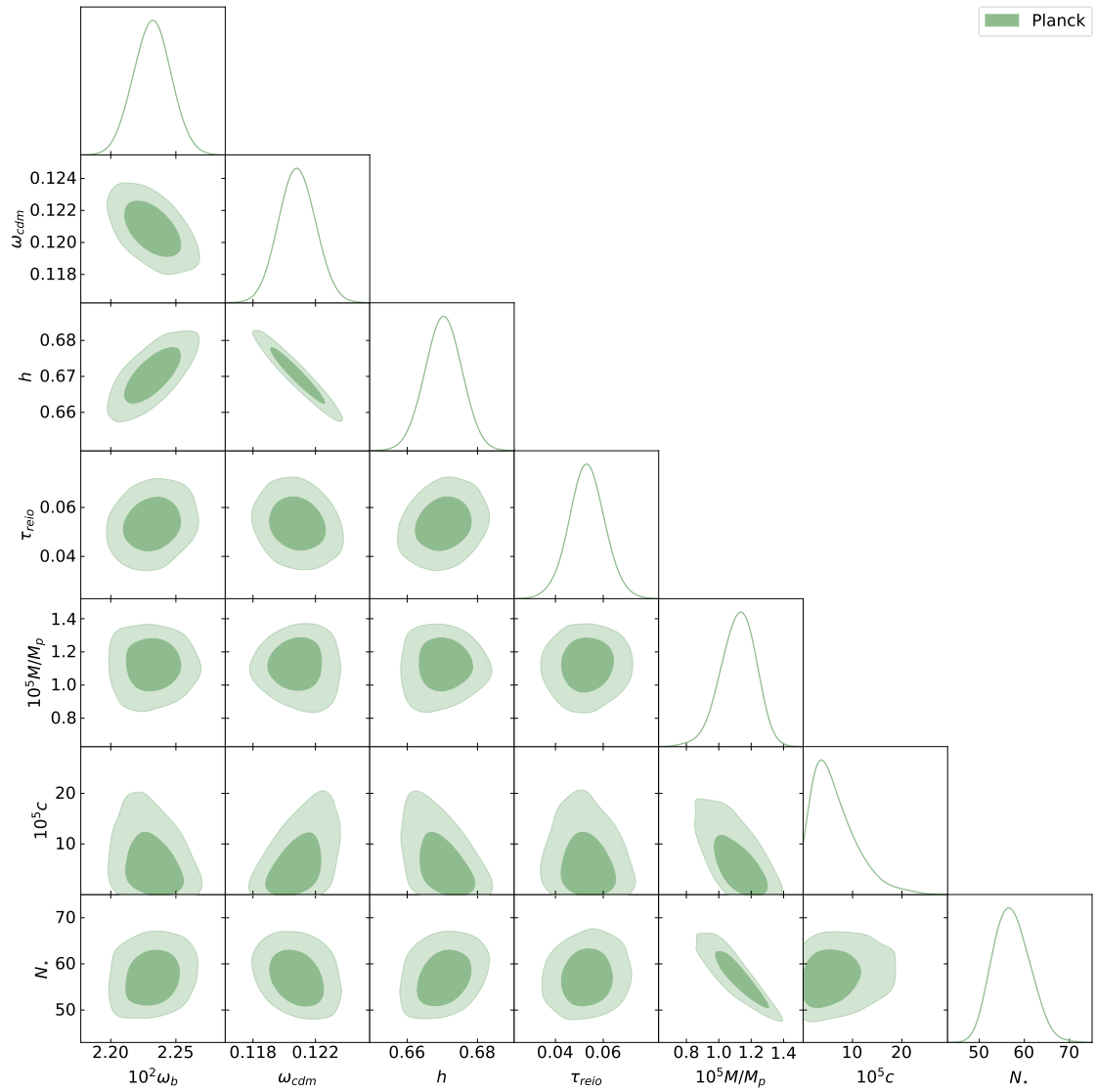


Figure 8.1: Marginalised posterior CMB posterior for the extended Starobinsky model, based on Planck ($TT, TE, EE + \text{low-}l$ $EE + \text{low-}l$ TT).

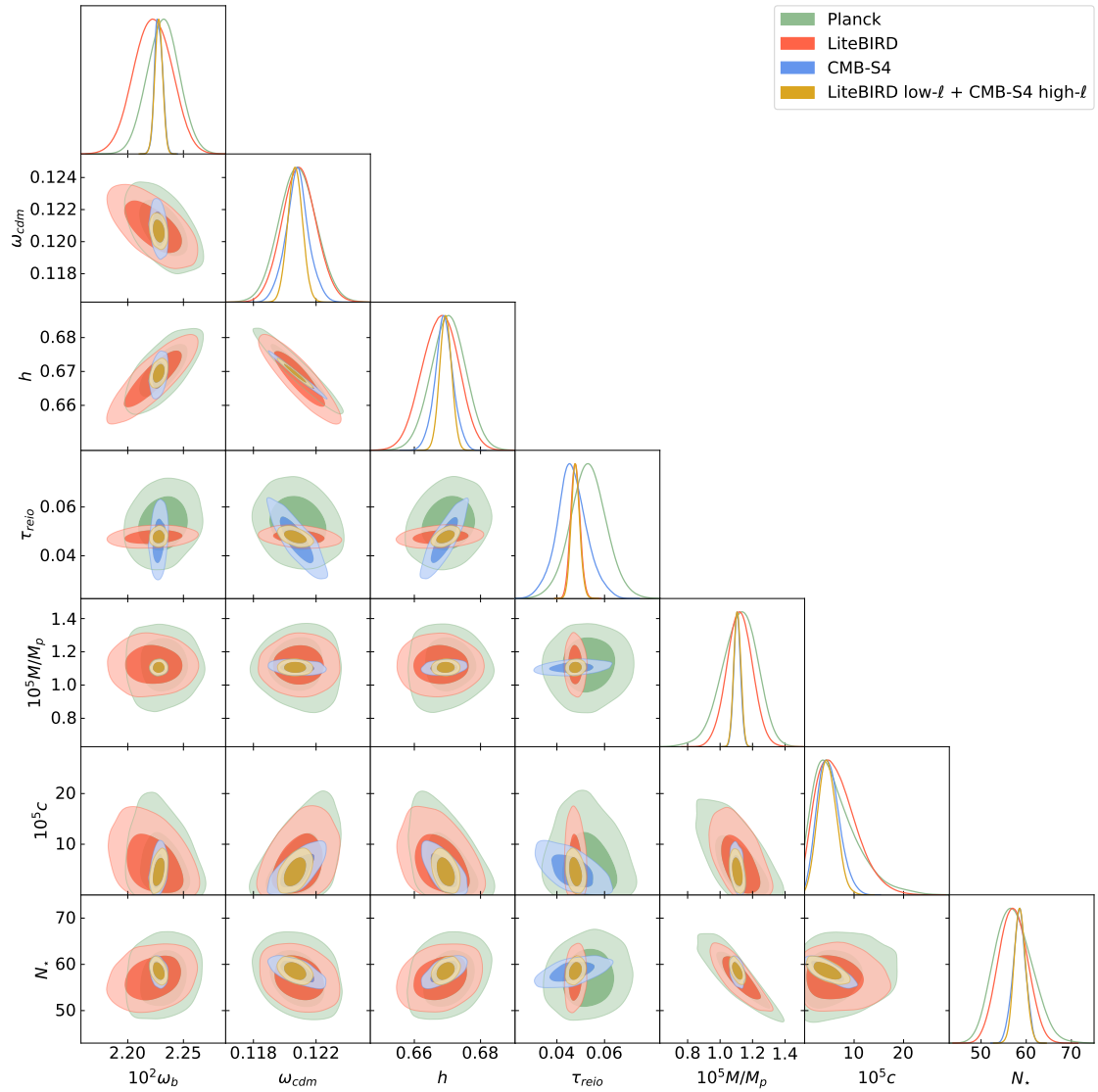


Figure 8.2: Marginalised CMB posterior for the extended Starobinsky model, based on Planck (TT , TE , EE +low- ℓ EE +low- ℓ TT), LiteBIRD, CMB-S4, and the consistent combination of LiteBIRD and CMB-S4.

8.3 Square Kilometer Array Forecasts

This section presents the sensitivity of the future 21cm experiment SKA to the extended Starobinsky parameters. For the analysis, the likelihood discussed in Sec. 7.5.2 is used. Without any information from the CMB, the power spectrum is not sufficient to constrain all the parameters of Eq. (8.1). However, adding information from the future CMB experiments LiteBIRD and CMB-S4 is more than enough to set tight constraints on the parameters. Even, solely a combination with the Planck data is sufficient to provide a measurement [48].

First, the addition of SKA to the two CMB experiments is investigated individually. The marginalised distributions are shown in Fig. 8.3. The contours show the same differences as the sole comparison of LiteBIRD and CMB-S4 in Fig. 8.2. It can be explained with the same properties of the experiments and how the parameters are determined, like in Sec. 8.2.

Finally, all of the experiments are combined, i.e. SKA with LiteBIRD low- ℓ and CMB-S4 high- ℓ , to give the best possible constraints. The corresponding best-fit, mean and 95%CL limits are given in Tab. 8.1 and the contours in Fig. 8.4. The cosmological parameters benefiting significantly from SKA are ω_{cdm} and h . Despite the fact, that this analysis is mainly focused on the inflationary parameters, this kind of improvement leads to a big advancement in the global analysis. While the combination with SKA still leaves a narrow correlation between N_{\star} and the Starobinsky parameter M , resulting in no improvement in the constraint for M

$$\frac{M}{M_P} = (1.064 \dots 1.148) \cdot 10^{-5} \quad (95\%CL), \quad (8.5)$$

it provides an improved reach in the second Starobinsky parameter c , as compared to the CMB projection of Eq. (8.4),

$$c = (2.89 \dots 5.73) \cdot 10^{-5} \quad (95\%CL). \quad (8.6)$$

The strong correlation between M and N_{\star} can be explained by looking at how CLASS solves the background equation (3.24). Given the number of e -folds of inflation, it evolves back in time until the initial field value is found, resulting in the desired N_{\star} . The initial value has a large influence on the exit point of the reference mode k_{\star} , which in turn defines the observables A_s and n_s , as they are just the amplitude and slope of the power spectrum evaluated at k_{\star} . In Sec. 5.4 the effect of the Starobinsky parameters M and c on these two observables is qualitatively investigated and it is shown that in the considered parameter space M mostly determines them. It might be possible to break this degeneracy by considering a more sophisticated reheating mechanism, which would set tighter constraints on the amount of inflation one gets.

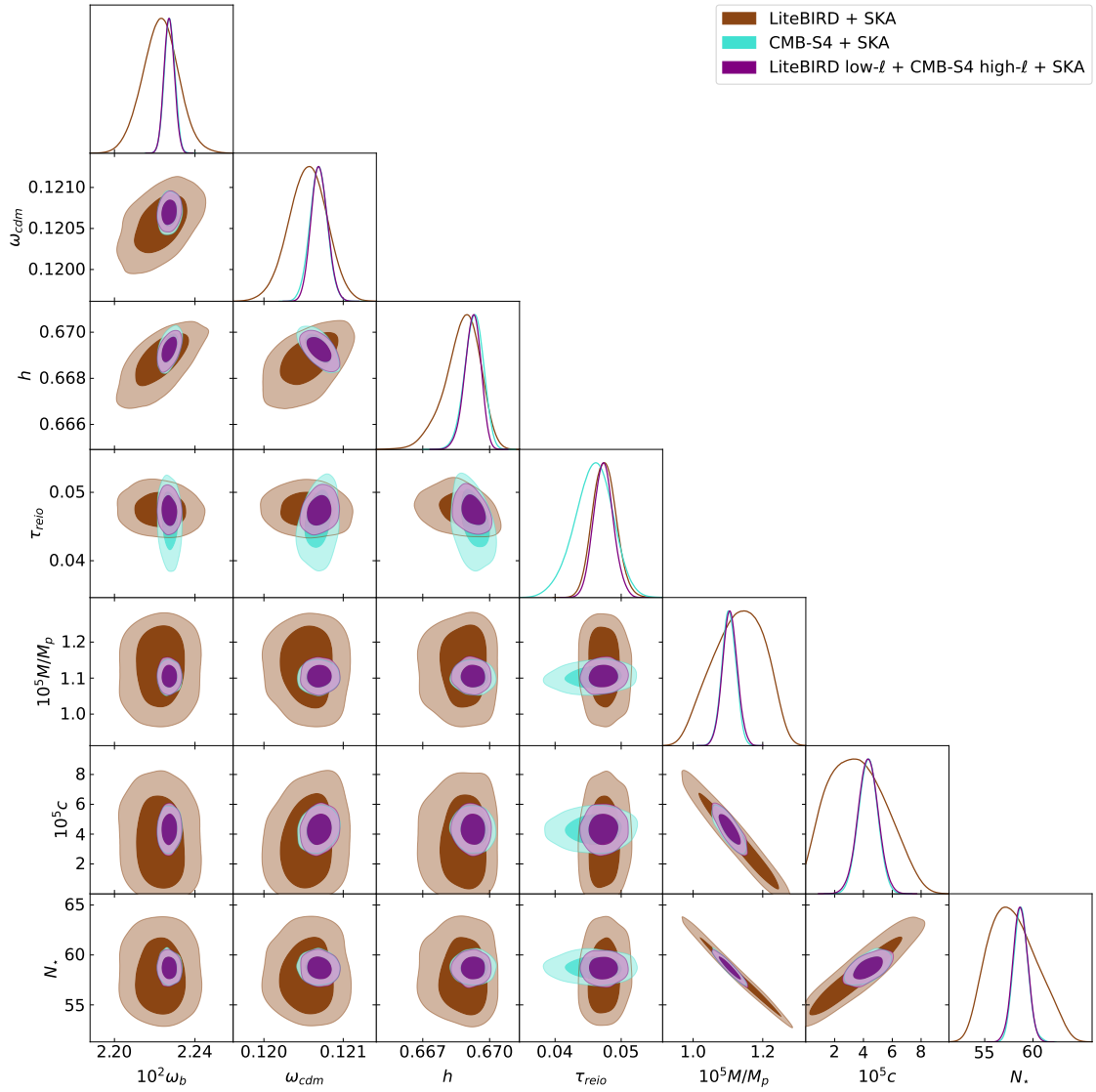


Figure 8.3: Marginalized CMB and SKA posteriors for the extended Starobinsky model, based on the individual combination of LiteBIRD and CMB-S4 with SKA.

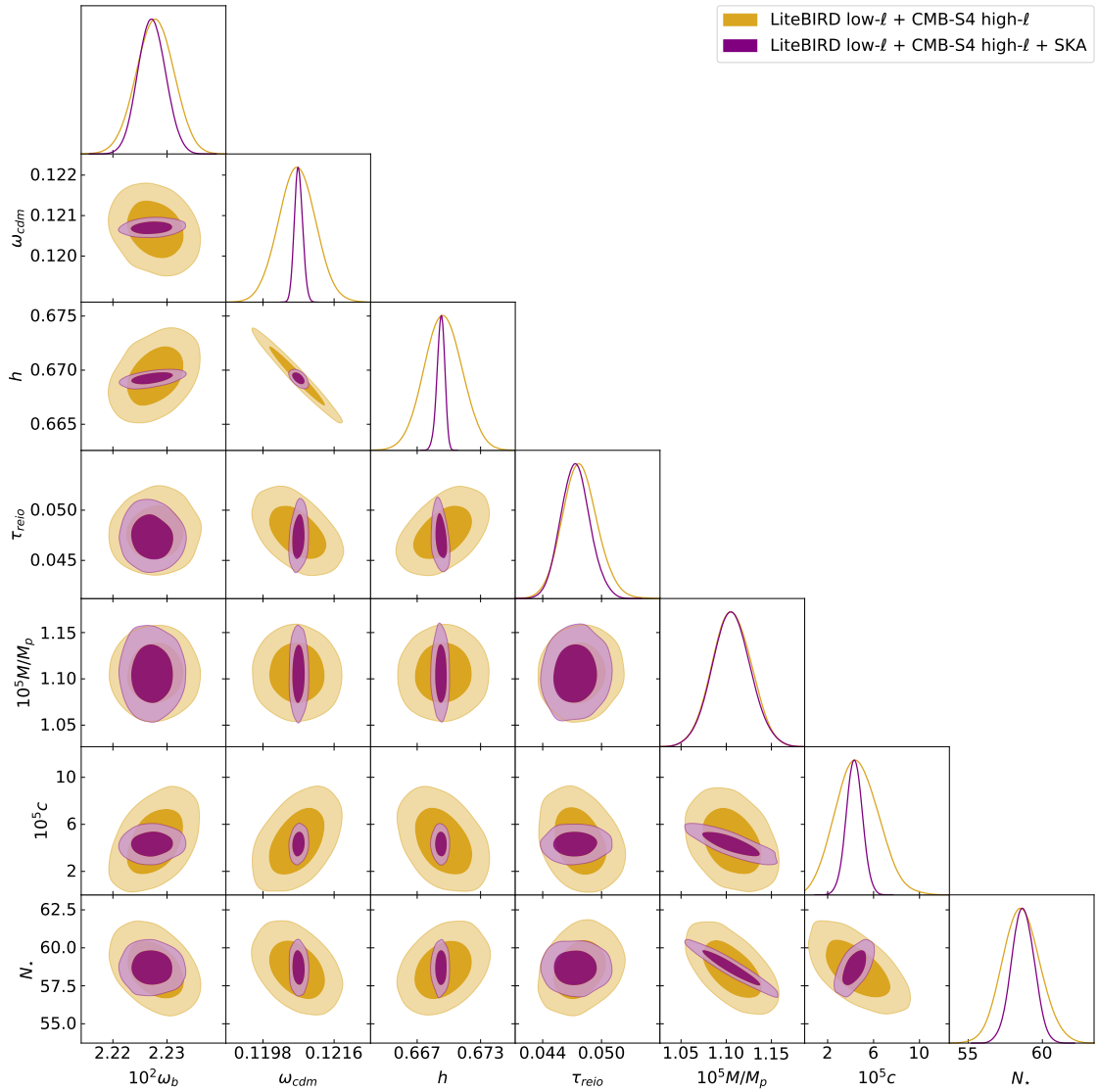


Figure 8.4: Marginalized CMB and SKA posteriors for the extended Starobinsky model, based on the combination of LiteBIRD low- ℓ and CMB-S4 with high- ℓ +SKA projections.

Data	Parameters	Best-fit	Mean $\pm\sigma$	95% lower	95% upper
Planck (TT, TE, EE +low- ℓ EE +low- ℓ TT)	100 ω_b	2.228	2.232 $^{+0.015}_{-0.015}$	2.203	2.26
	ω_{cdm}	0.1206	0.1208 $^{+0.0012}_{-0.0012}$	0.1185	0.1232
	h	0.6696	0.6703 $^{+0.0053}_{-0.0053}$	0.6600	0.6808
	τ_{reio}	0.04781	0.05315 $^{+0.0074}_{-0.0077}$	0.03764	0.0687
	$10^5 M/M_P$	1.103	1.119 $^{+0.117}_{-0.0987}$	0.9005	1.329
	$10^5 c$	4.135	6.069 $^{+2.840}_{-5.402}$	—	< 15.96
	N_*	58.24	57.17 $^{+3.73}_{-4.47}$	49.65	65.24
LiteBIRD	100 ω_b	2.229	2.223 $^{+0.018}_{-0.017}$	2.190	2.256
	ω_{cdm}	0.1204	0.1209 $^{+0.001}_{-0.0011}$	0.1188	0.1231
	h	0.6705	0.6679 $^{+0.0057}_{-0.0055}$	0.657	0.6785
	τ_{reio}	0.04735	0.04775 $^{+0.002}_{-0.002}$	0.04391	0.05171
	$10^5 M/M_P$	1.144	1.121 $^{+0.077}_{-0.077}$	0.9676	1.273
	$10^5 c$	2.633	6.345 $^{+2.996}_{-4.801}$	—	< 14.62
	N_*	57.79	57.08 $^{+3.18}_{-3.19}$	51.04	63.27
CMB-S4	100 ω_b	2.227	2.228 $^{+0.004}_{-0.004}$	2.221	2.235
	ω_{cdm}	0.121	0.1208 $^{+0.0007}_{-0.0007}$	0.1192	0.1223
	h	0.6681	0.669 $^{+0.0027}_{-0.0027}$	0.6634	0.6749
	τ_{reio}	0.04478	0.04634 $^{+0.0064}_{-0.0058}$	0.03258	0.05963
	$10^5 M/M_P$	1.098	1.105 $^{+0.021}_{-0.021}$	1.065	1.145
	$10^5 c$	5.166	4.794 $^{+1.923}_{-2.461}$	0.7769	9.543
	N_*	58.44	58.45 $^{+1.45}_{-1.35}$	55.66	61.26
LiteBIRD low- ℓ + CMB-S4 high- ℓ	100 ω_b	2.227	2.228 $^{+0.004}_{-0.004}$	2.221	2.235
	ω_{cdm}	0.1206	0.1207 $^{+0.0005}_{-0.0005}$	0.1197	0.1216
	h	0.6696	0.6695 $^{+0.0018}_{-0.0018}$	0.6659	0.673
	τ_{reio}	0.04829	0.04779 $^{+0.0017}_{-0.0019}$	0.04425	0.05148
	$10^5 M/M_P$	1.108	1.106 $^{+0.022}_{-0.021}$	1.064	1.147
	$10^5 c$	4.177	4.573 $^{+1.786}_{-1.944}$	1.015	8.300
	N_*	58.71	58.59 $^{+1.24}_{-1.25}$	56.15	61.08
LiteBIRD low- ℓ + CMB-S4 high- ℓ + SKA	100 ω_b	2.228	2.227 $^{+0.003}_{-0.003}$	2.222	2.232
	ω_{cdm}	0.1206	0.1207 $^{+0.0001}_{-0.0001}$	0.1205	0.1209
	h	0.6694	0.6692 $^{+0.0004}_{-0.0003}$	0.6685	0.670
	τ_{reio}	0.04792	0.04734 $^{+0.0014}_{-0.0016}$	0.04445	0.05033
	$10^5 M/M_P$	1.100	1.106 $^{+0.023}_{-0.023}$	1.064	1.148
	$10^5 c$	4.350	4.325 $^{+0.692}_{-0.690}$	2.891	5.734
	N_*	58.95	58.68 $^{+0.77}_{-0.75}$	57.20	60.18

Table 8.1: Best-fit values, mean, error bars, and 95%CL limits for the parameters shown in Fig. 8.1, Fig. 8.2 and Fig. 8.4.

CHAPTER 9

21cm Inference with Deep Learning

In this additional part of the thesis, some more technical advancements of the previous analysis are explored. However, the physical focus shifts from inflation to dark matter (DM), epoch of reionization (EoR) and cosmic dawn (CD) parameters. From one light-cone of 21cm brightness fluctuation $\Delta T_{21}(x, \nu)$, with on-sky coordinates x and frequency ν , one obtains the posterior for all parameters of interest, using an elaborate set-up of neural networks (NN), explicitly a 3D convolutional neural network (CNN) and a conditional invertible neural network (cINN). First, the production of the training set, then basics of deep learning are discussed. Furthermore, the exact architecture is presented before showing the results.

9.1 Dataset

To train a neural network one needs a large amount of training data. For this task, this means producing many light-cones (see Sec. 6.2) with the semi-numerical code 21cmFAST [49, 50]. They are produced from coeval cubes in redshift space by evolving density and velocity perturbations at first and second order perturbation theory using the Zel'dovich approximation [51].

The parameters of choice for this analysis are exactly like in Ref. [52]:

- Warm dark matter mass $m_{\text{WDM}} \in [0.3, 10]$ keV:
The limits are set very conservatively to observe a wide range of parameter behaviour, where at the upper limit the model looks almost like CDM.
- Dark matter density parameter $\Omega_{\text{m}} \in [0.2, 0.4]$:
It controls structure formation, where a large range is allowed including the limits set by Planck [4].
- Minimum virial temperature $T_{\text{vir}} \in [10^4, 10^{5.3}]$ K:
It determines the minimum virial temperature of halos for cooling to be efficient for star formation.
- Ionisation efficiency $\zeta \in [10, 250]$:
It is a composite parameter given by $\zeta = 30 \left(\frac{f_{\text{esc}}}{0.3}\right) \left(\frac{f_{\star}}{0.05}\right) \left(\frac{N_{\gamma}/b}{4000}\right) \left(\frac{2}{1+n_{\text{rec}}}\right)$, with

the fraction of ionising photons escaping into the IGM f_{esc} , the fraction of galactic gas in stars f_* , the number of ionising photons per baryon in stars $N_{\gamma/b}$ and the typical number density of recombinations for hydrogen in the IGM N_{rec} . The parameter range allows for a wide range of different recombination scenarios.

- Specific X-ray luminosity (with energies < 2 keV) per unit star formation rate that escapes host galaxies $L_X \in [10^{38}, 10^{42}] \text{ erg s}^{-1} M_{\odot} \text{ yr}$.
- X-ray energy threshold for self absorption by host galaxies $E_0 \in [100, 15000] \text{ eV}$: X-rays with energies below E_0 do not escape the host galaxy.

The light-cones are produced by sampling the parameters from a flat prior. The box size is given by 200 Mpc and a resolution of 1.42 Mpc. The redshift range is $z = 5 \dots 35$. Each light-cone has the dimension (2350, 140, 140) pixel, where it is important to note, that Ω_m has an impact on the length of each light-cone in terms of redshift. Hence, only a one with $\Omega_m = 0.4$ reaches $z = 35$. Due to the very large allowed parameter ranges some very unsensible combinations can happen. For that reason, it is imposed, that τ_{reio} is within 5σ of the Planck measurement (0.054 ± 0.007) [4] and the IGM mean neutral fraction at redshift 5 is below 0.1. With these settings 5000 light-cones are produced.

9.2 Basics of Deep Learning

This brief introduction to deep learning largely follows the lecture notes on *Modern Machine Learning for LHC Physicists*, given by Tilman Plehn at Heidelberg University [53].

9.2.1 Neural Network

A neural network can be understood as a fit function with a huge amount of parameters θ , such that

$$f_{\theta}(x) \approx f(x). \quad (9.1)$$

One can think of the very complex function $f_{\theta}(x)$ in terms of small building blocks, which are all connected to each other. The complexity of the problem defines how many building blocks one needs. The network consists of multiple layers, where each output is the input of the next one:

$$x \rightarrow x^{(1)} \rightarrow x^{(2)} \rightarrow \dots \rightarrow x^{(N)} \equiv f_{\theta}(x), \quad (9.2)$$

if the network is fully connected, or dense, the next vector collects all information from the previous vector. Splitting the vector x in each D entries each step can be written as

$$x_i^{(n)} = W_{ij}^{(n)} x_j^{(n-1)} + b_i^{(n)}, \quad (9.3)$$

where b is called the bias and W the weights. For now, this can only describe linear functions, but of course one needs non-linearity. One very simple way to introduce

non-linearity in the network is by an activation function on each node. A common choice is the rectified linear unit

$$\text{ReLU}(x_j) \equiv \max(0, x_j) = \begin{cases} 0 & x_j \leq 0 \\ x_j & x_j > 0 \end{cases}, \quad (9.4)$$

which alters Eq. (9.3) to

$$x_i^{(n)} = \text{ReLU}[W_{ij}^{(n)} x_j^{(n-1)} + b_i^{(n)}]. \quad (9.5)$$

Now, the network can describe very complex non-linear function and by adding more layers it can describe even more complex functions.

To determine the parameters θ one minimizes a loss function \mathcal{L} , which can be any sensible function, depending on the situation. To illustrate the idea, a simple mean-squared-error (MSE) loss is chosen. It is given by

$$\mathcal{L} \propto \sum_j |f_j - f_\theta(x_j)|^2 \quad (9.6)$$

and to minimise it one needs the derivative. To not overcomplicate notation, the bias is set to zero for now and one gets

$$\frac{d\mathcal{L}}{dW_{ij}^{(n)}} = -2\sqrt{\mathcal{L}} (W^{(N)} \dots W^{(n+1)})_{ij} x_j^{(n-1)}. \quad (9.7)$$

The derivatives of the network are simply given by the chain rule. An efficient algorithm to compute the derivatives for each layer is called *back propagation*. To find the minimum of the loss, one needs to update the network parameters iteratively. One choice for this procedure is called stochastic gradient descent method, where the update happens in the following way

$$\theta_j^{(t+1)} = \theta_j^{(t)} - \alpha \left\langle \frac{\partial \mathcal{L}^{(t)}}{\partial \theta_j} \right\rangle \quad \text{with} \quad \theta_j \in \{b, W\}, \quad (9.8)$$

where expectation value is defined over a minibatch, as it is to computationally expensive to take it over the whole training set. The learning rate α determines the size of the step one takes in the direction of the minimum. Of course, more sophisticated version of this update procedure exist, the most prominent one being the Adam optimizer [54].

9.2.2 Convolutional Neural Network

The convolutional neural network (CNN) is an idea from research on image recognition, which is essentially what a light-cone is, a 3D image. A CNN learns the correlation between pixels, by performing a convolution with a matrix-like filter encoding information about the neighbourhood of each pixel. There exist a few standard operations for CNNs, they will be explained for a 2D CNN, due do the fact that one can better imagine them and they are easily generalised to higher dimension. The network input is a 2D image as $(n \times n)$ -dimensional matrix and the output is usually a 1D vector, where the size depends on the task.

- Zero padding $(n \times n) \rightarrow (n + 2 \times n + 2)$: It increases the size of the input by adding zeros at the boundary to use a filter for the pixels at the boundary.
- Convolution $(n \times n) \rightarrow (n - n_{\text{c-size}} + 1 \times n - n_{\text{c-size}} + 1)$: The input image is convoluted with learnable filters of size $n_{\text{c-size}} \times n_{\text{c-size}}$ and a stride of one. They play the role of nodes

$$x'_{ij} = \text{ReLU} \sum_{r,s} [W_{rs} x_{i+r,j+s} + b]. \quad (9.9)$$

- Feature maps $n_{f\text{-maps}} \times (n \times n) \rightarrow n_{f\text{-maps}} \times (n \times n)$: To increase the expressibility of the network multiple filters are introduced. They turn one image into $n_{f\text{-maps}}$ feature maps $x^{(k)}$ which mixes information from all input maps

$$x^{(k)} = \sum_{l=0}^{n_{f\text{-maps}}-1} \sum_{r,s} [W_{rs}^{(kl)} x_{i+r,j+s}^{(l)} + b^{(k)}]. \quad (9.10)$$

This defines a convolutional layer, of which multiple are stacked.

- Pooling: $(n \times n) \rightarrow (n/p \times n/p)$: The size of a feature map is reduced by a downsampling algorithm. A patch with fixed size $p \times p$ is reduced to single pixel by taking the average or the maximum. This is needed, as one wants a compact network output in the end.
- Flattening: $(n \times n) \rightarrow (n^2 \times 1)$: The final output of the network has to be a vector for this task. Therefore, the 2D object is flattened to a vector.
- Fully connected layers $n^2 \rightarrow n_{\text{d-node}}$: On the flattened vector a normal fully connected NN can be used like in Eq. (9.5), where $n_{\text{d-node}}$ depends on the task at hand. Multiple of dense layers are stacked at the end before giving the final output.

Training the CNN works exactly like for a standard NN.

9.3 BayesFlow

A CNN is already a great tool for parameter recovery. However, one only gets an estimate without an error. For Bayesian inference tasks, the BayesFlow set-up [55] is a good architecture. This section presents how the combination of a summary network and a conditional invertible neural network (cINN) can learn the posterior of a given parameter set, following Ref. [55]. The idea is illustrated in Fig. 9.1, where the networks are trained with simulations from a model of interest. Then, only one measurement is needed to obtain the posterior for the parameters of the model.

First, one has to deal with notation. The number of parameters of a mathematical model are denoted as D and the parameters themselves as a vector $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_D)$, the number of observations in a dataset as N . The simulated data is $\mathbf{x}_{1:N} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)$ with \mathbf{x}_i a vector or scalar. Observed or test data gets a superscript o , such that it reads $\mathbf{x}_{1:N}^o$. The trainable parameters of the summary neural network are ψ and the one from invertible neural network are ϕ .

9.3.1 Finding the Posterior

For now, assume that $N = 1$ to not make the notation too confusing and that the summary network does not exist, i.e. the raw data \mathbf{x} is fed directly into the invertible neural network. The goal is to find an approximate posterior:

$$p_\phi(\boldsymbol{\theta}|\mathbf{x}) \approx p(\boldsymbol{\theta}|\mathbf{x}) \quad (9.11)$$

for all possible $\boldsymbol{\theta}$ and \mathbf{x} . Thus, the objective is to find network parameters $\hat{\phi}$, which minimize the Kullback-Leibler (KL) divergence between the approximate and true posterior

$$\hat{\phi} = \arg \min_{\phi} \mathbb{E}_{p(\mathbf{x})} [\text{KL}(p(\boldsymbol{\theta}|\mathbf{x}) || p_\phi(\boldsymbol{\theta}|\mathbf{x}))] \quad (9.12)$$

$$= \arg \min_{\phi} \mathbb{E}_{p(\mathbf{x})} [\mathbb{E}_{p(\boldsymbol{\theta}|\mathbf{x})} (\log p(\boldsymbol{\theta}|\mathbf{x}) - \log p_\phi(\boldsymbol{\theta}|\mathbf{x}))] \quad (9.13)$$

$$= \arg \max_{\phi} \mathbb{E}_{p(\mathbf{x})} [\mathbb{E}_{p(\boldsymbol{\theta}|\mathbf{x})} [\log p_\phi(\boldsymbol{\theta}|\mathbf{x})]] \quad (9.14)$$

$$= \arg \max_{\phi} \int \int p(\mathbf{x}, \boldsymbol{\theta}) \log p_\phi(\boldsymbol{\theta}|\mathbf{x}) d\mathbf{x} d\boldsymbol{\theta}, \quad (9.15)$$

where the exact posterior density can be dropped as it does not depend on ϕ and the minus sign changes arg min to arg max.

The approximate posterior can be reparameterized with a conditional invertible neural network (cINN) $f_\phi : \mathbb{R} \rightarrow \mathbb{R}$ for which f_ϕ^{-1} exists. This network implements a normalizing flow between $\boldsymbol{\theta}$ and a Gaussian latent variable \mathbf{z} :

$$\boldsymbol{\theta} \sim p_\phi(\boldsymbol{\theta}|\mathbf{x}) \iff \boldsymbol{\theta} = f_\phi^{-1}(\mathbf{z}; \mathbf{x}) \text{ with } \mathbf{z} \sim \mathcal{N}_D(\mathbf{z}|\mathbf{0}, I) \quad (9.16)$$

allowing the use of change of variable rule yielding

$$p_\phi(\boldsymbol{\theta}|\mathbf{x}) = p(\mathbf{z} = f_\phi(\boldsymbol{\theta}; \mathbf{x})) \left| \det \left(\frac{\partial f_\phi(\boldsymbol{\theta}; \mathbf{x})}{\partial \boldsymbol{\theta}} \right) \right|. \quad (9.17)$$

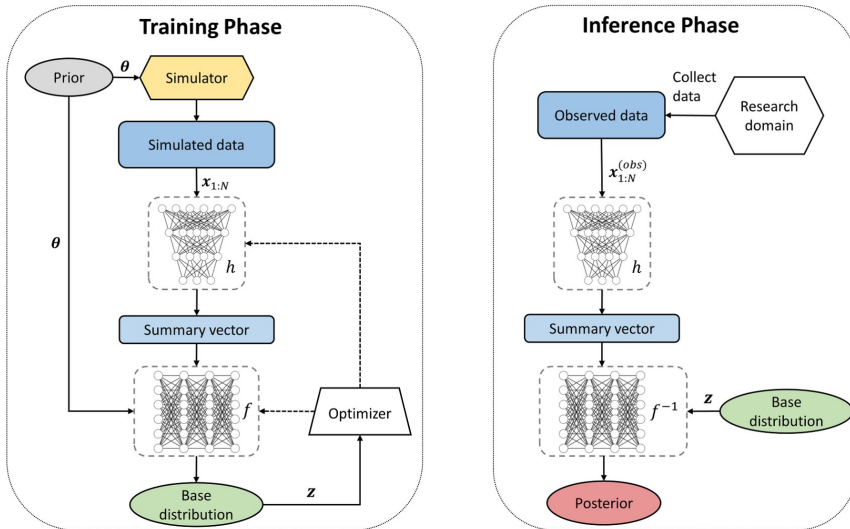


Figure 9.1: Conceptual idea of performing Bayesian inference with BayesFlow. Figure from Ref. [55].

Therefore, Eq. (9.15) becomes

$$\hat{\phi} = \arg \max_{\phi} \int \int p(\mathbf{x}, \boldsymbol{\theta}) (\log p(f_{\phi}(\boldsymbol{\theta}; \mathbf{x})) + \log |\det \mathbf{J}_{f_{\phi}}|) d\mathbf{x} d\boldsymbol{\theta}, \quad (9.18)$$

with the Jacobian $\mathbf{J}_{f_{\phi}}$ of f_{ϕ} evaluated at $\boldsymbol{\theta}$ and \mathbf{x} .

By using the a Monte Carlo estimate it is possible to maximize the likelihood with M simulated datasets and the data-generating parameters $\{(\mathbf{x}^{(m)}, \boldsymbol{\theta}^{(m)})\}_{m=1}^M$:

$$\hat{\phi} = \arg \min_{\phi} \frac{1}{M} \sum_{m=1}^M -\log p(f_{\phi}(\boldsymbol{\theta}^{(m)}; \mathbf{x}^{(m)})) - \log |\det \mathbf{J}_{f_{\phi}}^{(m)}| \quad (9.19)$$

$$= \arg \min_{\phi} \frac{1}{M} \sum_{m=1}^M \left(\frac{\|f_{\phi}(\boldsymbol{\theta}^{(m)}; \mathbf{x}^{(m)})\|_2^2}{2} - \log |\det \mathbf{J}_{f_{\phi}}^{(m)}| \right), \quad (9.20)$$

where the first term comes from the fact, that \mathbf{z} is prescribed to a unit Gaussian distribution.

Now, it is time to reintroduce the summary network with its output vector $\tilde{\mathbf{x}} = h_{\psi}(\mathbf{x}_{1:N})$. The summary network is trained jointly with the cINN, which changes the objective to

$$\hat{\phi}, \hat{\psi} = \arg \max_{\phi, \psi} \mathbb{E}_{p(\mathbf{x}, \boldsymbol{\theta}, N)} [\log p_{\phi}(\boldsymbol{\theta} | h_{\psi}(\mathbf{x}_{1:N}))], \quad (9.21)$$

and the Monte Carlo estimate to

$$\hat{\phi}, \hat{\psi} = \arg \min_{\phi, \psi} \frac{1}{M} \sum_{m=1}^M \left(\frac{\|f_{\phi}(\boldsymbol{\theta}^{(m)}; h_{\psi}(\mathbf{x}_{1:N}^{(m)}))\|_2^2}{2} - \log |\det \mathbf{J}_{f_{\phi}}^{(m)}| \right). \quad (9.22)$$

Eq. (9.22) is going to be the loss function \mathcal{L} and is therefore minimized in training yielding the correct posterior after successfully finding the minimum. The minimization can be achieved with any gradient descend algorithm.

9.3.2 Composing Invertible Networks

The cINN is composed of multiple affine building blocks (ABC) [56]. An ABC performs an invertible non-linear transformation. Therefore, one not only get $f_{\phi}(\mathbf{u}) = \mathbf{v}$ but also $f_{\phi}^{-1}(\mathbf{v}) = \mathbf{u}$ for free. Here, \mathbf{u} is the input vector and \mathbf{v} the output vector. To make the transformation trivially invertible the input vector is split into two parts

$$\mathbf{u} = (\mathbf{u}_1, \mathbf{u}_2) \text{ with } \mathbf{u}_1 = u_{1:D/2} \text{ and } \mathbf{u}_2 = u_{D/2+1:D}, \quad (9.23)$$

with the $D/2$ is to understood as a floor division. The operation of the ABC is now the following:

$$\mathbf{v}_1 = \mathbf{u}_1 \odot \exp(s_1(\mathbf{u}_2)) + t_1(\mathbf{u}_2) \quad (9.24)$$

$$\mathbf{v}_2 = \mathbf{u}_2 \odot \exp(s_2(\mathbf{v}_1)) + t_1(\mathbf{v}_1), \quad (9.25)$$

with the element-wise multiplication \odot and the outputs are again concatenated. The inverse operation is given by:

$$\mathbf{u}_2 = (\mathbf{v}_2 - t_2(\mathbf{v}_1)) \odot \exp(-s_2(\mathbf{v}_1)) \quad (9.26)$$

$$\mathbf{u}_1 = (\mathbf{v}_1 - t_1(\mathbf{u}_2)) \odot \exp(-s_1(\mathbf{u}_2)). \quad (9.27)$$

The functions s_1, s_2, t_1, t_2 are four separate fully connected neural networks, in this case with a ReLU activation. These networks do not need to be invertible for the cINN to be invertible, as they are always only used in the forward direction. Due to the construction of the transformation, the Jacobian is always a strictly upper or lower triangular matrix, which make the computation of the determinant very easy. For the whole cINN multiple of these blocks are put together, where the output of one is the input of the next, but the dimensions are permuted to increase the expressivity.

So far the output of the summary network is not incorporated in the cINN. This is achieved very easily by giving its output vector $\tilde{\mathbf{x}}$ as an additional input at each block, i.e.

$$\mathbf{v}_1 = \mathbf{u}_1 \odot \exp(s_1(\mathbf{u}_2, \tilde{\mathbf{x}})) + t_1(\mathbf{u}_2, \tilde{\mathbf{x}}) \quad (9.28)$$

$$\mathbf{v}_2 = \mathbf{u}_2 \odot \exp(s_2(\mathbf{v}_1, \tilde{\mathbf{x}})) + t_2(\mathbf{v}_1, \tilde{\mathbf{x}}). \quad (9.29)$$

So a complete pass through the cINN can expressed as $f_\phi(\boldsymbol{\theta}; \tilde{\mathbf{x}}) = \mathbf{z}$ and the inverse direction $f_\phi^{-1}(\mathbf{z}; \tilde{\mathbf{x}}) = \boldsymbol{\theta}$.

9.4 Architecture and Training

9.4.1 Architecture

The set-up consists of two neural networks, where the summary network is far more expressive compared to the cINN. As a summary network a 3D CNN is used, following Ref. [52] and their use for parameter inference. Its exact architecture is given in Tab. 9.1. The kernel size of the first filter has an asymmetric shape of $(3 \times 3 \times 102)$ to reflect the difference between fluctuations in z -direction and spatial-direction. Furthermore, a stride of $(1 \times 1 \times 102)$ reduces the dimensionality of the following layers, making it computationally cheaper, while still capturing the relevant physics. The kernel size of the hidden layers is set to $(3 \times 3 \times 2)$ and max pooling layers are applied only in the spatial direction. In front of three fully connected layers is one global average pooling layer to impede overfitting, i.e. learning the training data so well, that the network cannot interpolate to the test data. The hidden layers use a ReLU activation function.

The cINN is made up of the building blocks, described in Sec. 9.3.2. Here, 8 ABCs are set in sequence where each of the internal fully connected NN has 256 nodes with one hidden layer and ReLU activation.

For composing the cINN FrEIA (Framework for Easily Invertible Architectures) [57] is used. While the training and evaluation is performed in the PyTorch framework [58].

Layer	Shape
Input Layer	(1,140,140,2350)
3x3x102 Conv3D	(32,138,138,23)
3x3x2 Conv3D	(32,136,136,22)
2x2x1 Max Pooling	(32,68,68,22)
3x3x2 Conv3D	(64,66,66,21)
1x1x0 Zero Padding	(64,66,66,20)
3x3x2 Conv3D	(64,66,66,20)
2x2x1 Max Pooling	(64,33,33,20)
3x3x2 Conv3D	(128,31,31,19)
1x1x0 Zero Padding	(128,33,33,19)
3x3x2 Conv3D	(128,31,31,18)
Global Average Pooling	(128)
Dense	(128)
Dense	(128)
Dense	(128)
Dense	(6)

Table 9.1: 3D CNN model summary.

9.4.2 Training

The dataset of 5000 light-cones is divided into 1000 for testing, 400 for validation and the rest for training. The batch size is always set to 8 and for adjusting the weight, the Adam optimizer is used. Training the full set-up is composed of three stages:

Stage 1:

First, only the summary network is trained. As a loss function a normal MSE-loss is used, since the correct labels of the parameters are known. The learning rate is initially set to $4 \cdot 10^{-4}$ and after 15 and 20 epochs decreased by a factor of 0.5. The network is trained for 31 epochs and the evolution of the loss can be seen in Fig. 9.2a. Note, that one does not need perfect convergence here, as this is only the first stage.

Stage 2:

Next, only the cINN is trained with the fixed, pretrained CNN from stage 1. The loss is the one defined in Eq. (9.20) and its evolution is shown in Fig.9.2b. For this training a constant learning rate of $4 \cdot 10^{-4}$ is used and training is stopped after 150 epochs.

Stage 3:

Finally, everything is trained together, which is necessary to ensure convergence to the correct minimum. The loss is defined in Eq. (9.22). The learning rate is adjusted every five steps. Starting from $4 \cdot 10^{-4}$ it is reduced by a factor of 0.3 until it stays constant after epoch 20. Its effect can be seen in the loss curve. The training is aborted after 32 epochs.

This alternating training scheme is necessary, because the summary network is much more complicated than the cINN and needs a head start in training. Otherwise, the

summary network makes random predictions and it never learns a useful summary of the input. The architecture and hyperparameters are summarized in Tab. 9.2.

Number of parameters	6
Batch size	8
Batches per epoch	450
Output dimension summary network	6
Coupling layers	8
Fully connected coupling layer architecture	256, for all layers
Epochs	31,150,32
Training / testing points	3600 / 1000

Table 9.2: Network set-up and hyperparameters.

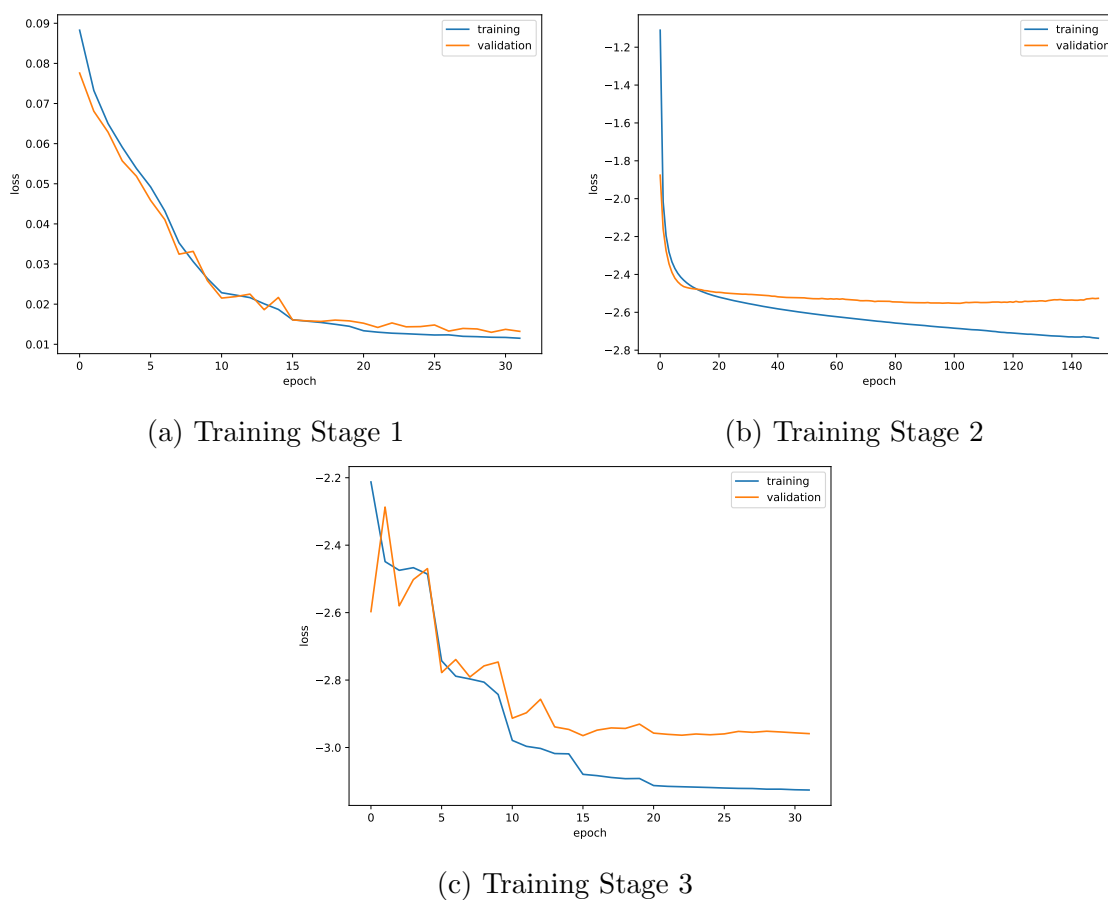


Figure 9.2: MSE-loss for the training and validation data set as a function of the training epoch.

9.4.3 Performance Validation

To assess the performance of the set-up a number of metrics are used:

- **Parameter Recovery:**

The mean of the marginalized posterior should statistically frequently coincide with the true value of the parameter, for a perfect approximation. To quantify the deviation, two metrics are used, which are based on how sample of true parameters $\{\theta^{(m)}\}_{m=1}^M$ is explained by a sample of estimated parameter $\{\hat{\theta}^{(m)}\}_{m=1}^M$. The estimated parameters are the means.

Coefficient of determination (R^2): It measures the proportion of variance and the formula is given by

$$R^2 = 1 - \sum_{m=1}^M \frac{(\theta^{(m)} - \hat{\theta}^{(m)})^2}{(\theta^{(m)} - \bar{\theta}^{(m)})^2}. \quad (9.30)$$

For perfect parameter recovery R^2 equals 1.

Normalized root mean square deviation (NRMSE): It is given by

$$\text{NRMSE} = \frac{\sqrt{\frac{1}{M} \sum_{m=1}^M (\theta^{(m)} - \hat{\theta}^{(m)})^2}}{\theta_{\max} - \theta_{\min}}, \quad (9.31)$$

where the normalisation allows for comparison throughout all parameters.

- **Simulation-Based Calibration [59]:**

It is a self-consistency check to visually detect systematic biases. One generates a histogram using the rank statistic

$$r^{(m)} = \sum_{l=1}^L I_{[\theta^{(l)} < \tilde{\theta}^{(m)}]}, \quad (9.32)$$

where $\tilde{\theta}^{(m)}$ is a true parameter, used to generate data, and $\theta^{(l)}$ are samples from the approximate posterior for the previously generated data. If it is self-consistent the resulting histogram is uniformly distributed. Common failure modes, like over- and under-confidence as well as biases are visually detectable.

- **Latent Space examination:**

The training objective is to map to unit Gaussian in the latent space. Therefore, upon good convergence, one expects the latent space to have a Gaussian structure.

- **Calibration Error [60]:**

Err_{cal} quantifies how well the coverage of an approximate posterior matches the coverage of an unknown true posterior. For each parameter the marginalized approximate posterior is given and the α -credible intervals can be calculated, with $\alpha \in (0, 1)$. For each α the fraction of true parameter values lying in the interval is denoted with α_{θ} . For a perfectly calibrated approximate posterior: $\alpha_{\theta} = \alpha$. The calibration error is defined as

$$Err_{\text{cal}} = \sum_{l=1}^L |\alpha_{\theta}^{(l)} - \alpha^{(l)}|, \quad (9.33)$$

with L equally spaced $\alpha^{(l)}$. The calibration error ranges from 0 to 1, with 0 indicating perfect calibration.

9.5 Results

This section discusses the results of parameter inference with the set-up and training described in Sec. 9.4. First, in Fig. 9.3 the parameter recovery is shown. There are clear differences in performance, listed in Tab. 9.3.

The worst parameter is m_{WDM} , especially for large masses. But for that, there is a physical explanation, as for large m_{WDM} , it starts to look more like CDM. However, the networks uncertainty accounts for that, as there is actually information lost and an analysis with power spectra comes to the same result [61]. The outliers in T_{vir} can be explained with m_{WDM} as well [52]. A threshold for early star formation is set by T_{vir} , but also by $M_{\text{Jeans}} \propto (\Omega_{\text{m}} h^2)^{1/2} \left(\frac{m_{\text{WDM}}}{\text{keV}}\right)^{-4} M_{\odot}$. For large m_{WDM} the Jeans mass limit is dominating and the minimum virial temperature has little effect on the era of reionization, resulting in the outliers. The errorband is dominated by those, for small m_{WDM} the network is very certain about its output. The R^2 -value of E_0 is not perfect, but the uncertainty is larger as well. The true value is therefore still within the one σ region. The rest of the parameter set shows almost perfect parameter recovery.

In order to check for consistency, a simulation based calibration is performed. The resulting histograms are shown in Fig. 9.4, for a sample size of 10000 and 50 bins. The grey shaded region is the expected variation, based on the 99% quantile of a Binomial. Therefore, one has the expected variance and the histograms are uniformly distributed, implying self-consistent sampling.

As a last performance check, the latent space is investigated. As the learning objective is to map to normal Gaussian, for a converged network one would expect it to be distributed as such. It is shown in Fig. 9.5. The distribution follows a Gaussian, but for some parameters is a slightly wider. However, it is safe to state, that the network is in a minimum.

	m_{WDM}	Ω_{m}	E_0	L_X	T_{vir}	ζ
R^2	0.621	0.981	0.803	0.987	0.764	0.969
NRMSE	0.181	0.039	0.130	0.032	0.135	0.048
Err_{cal}	0.011	0.007	0.025	0.016	0.050	0.028

Table 9.3: Various metrics (see Sec. 9.4.3) to check the performance of the network for all parameters.

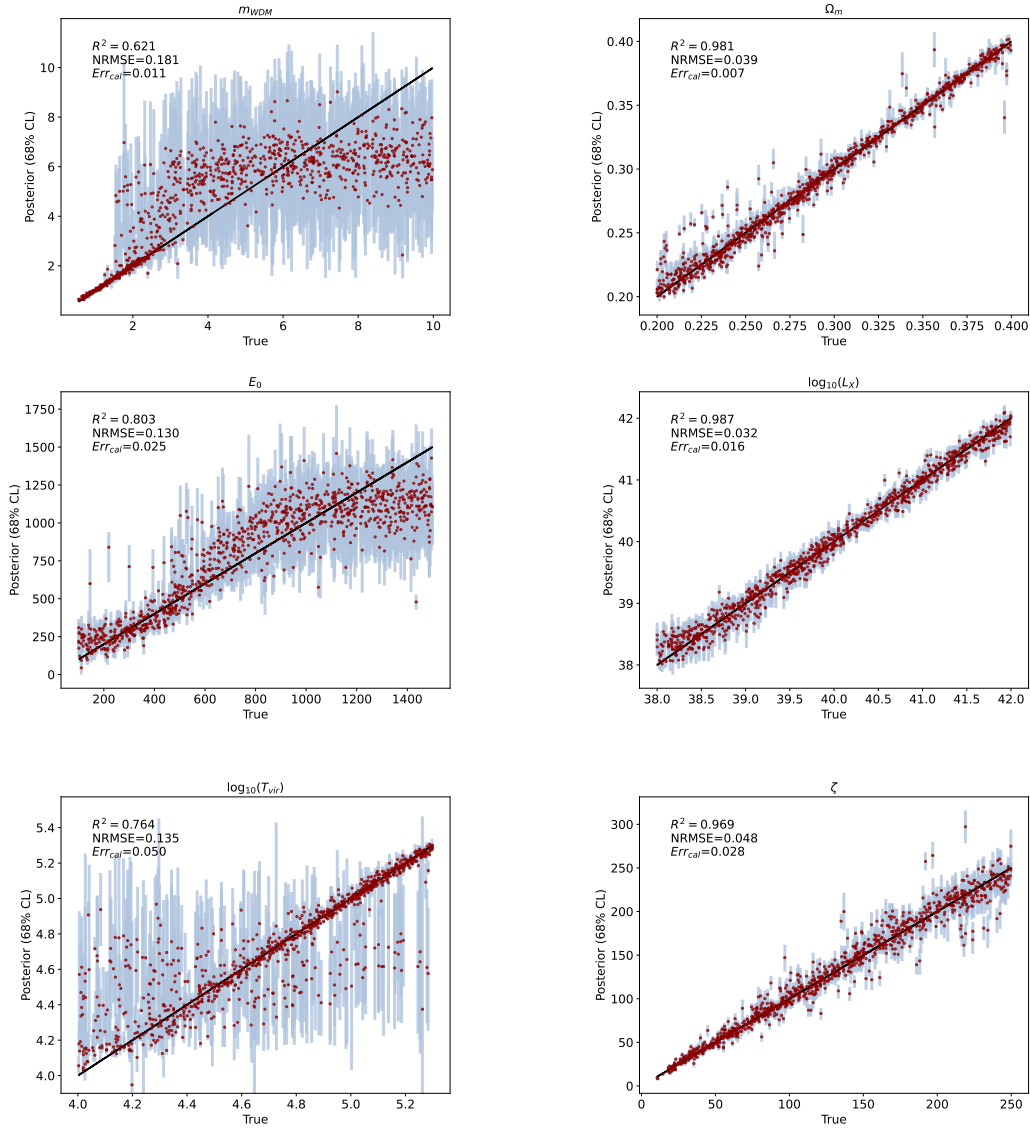


Figure 9.3: Parameter recovery, with the mean of the marginalized posterior (red dots) and the 68% Cl intervals in light blue. R^2 , NRMSE and Err_{cal} are calculated according to Eq. (9.30), Eq. (9.31) and Eq. (9.33), respectively.

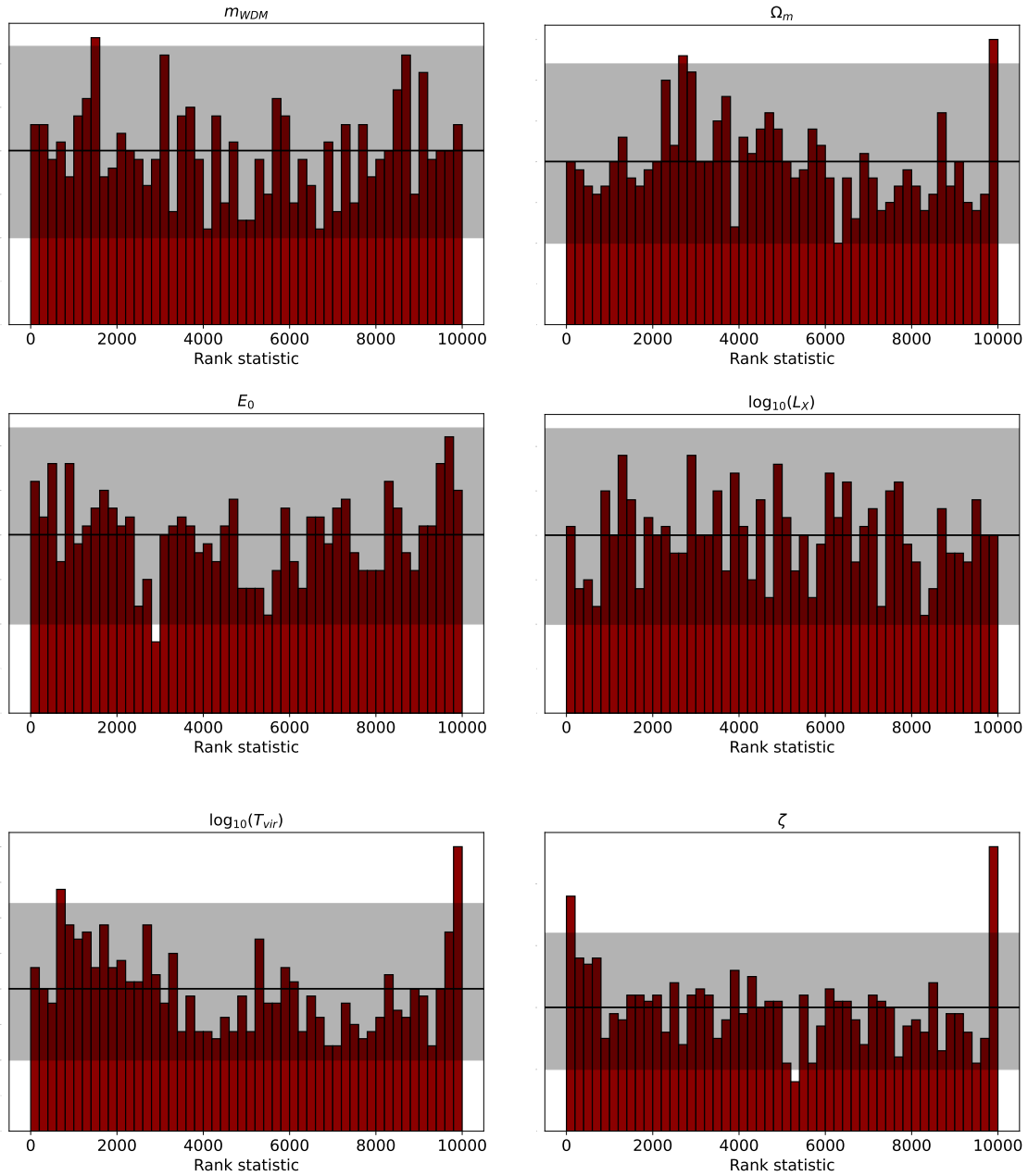


Figure 9.4: Simulation based calibration for all parameters. A uniform distribution indicates no bias, with the shaded region indicating the variation one expects in a uniform distribution, here the 99% quantile of a Binomial.

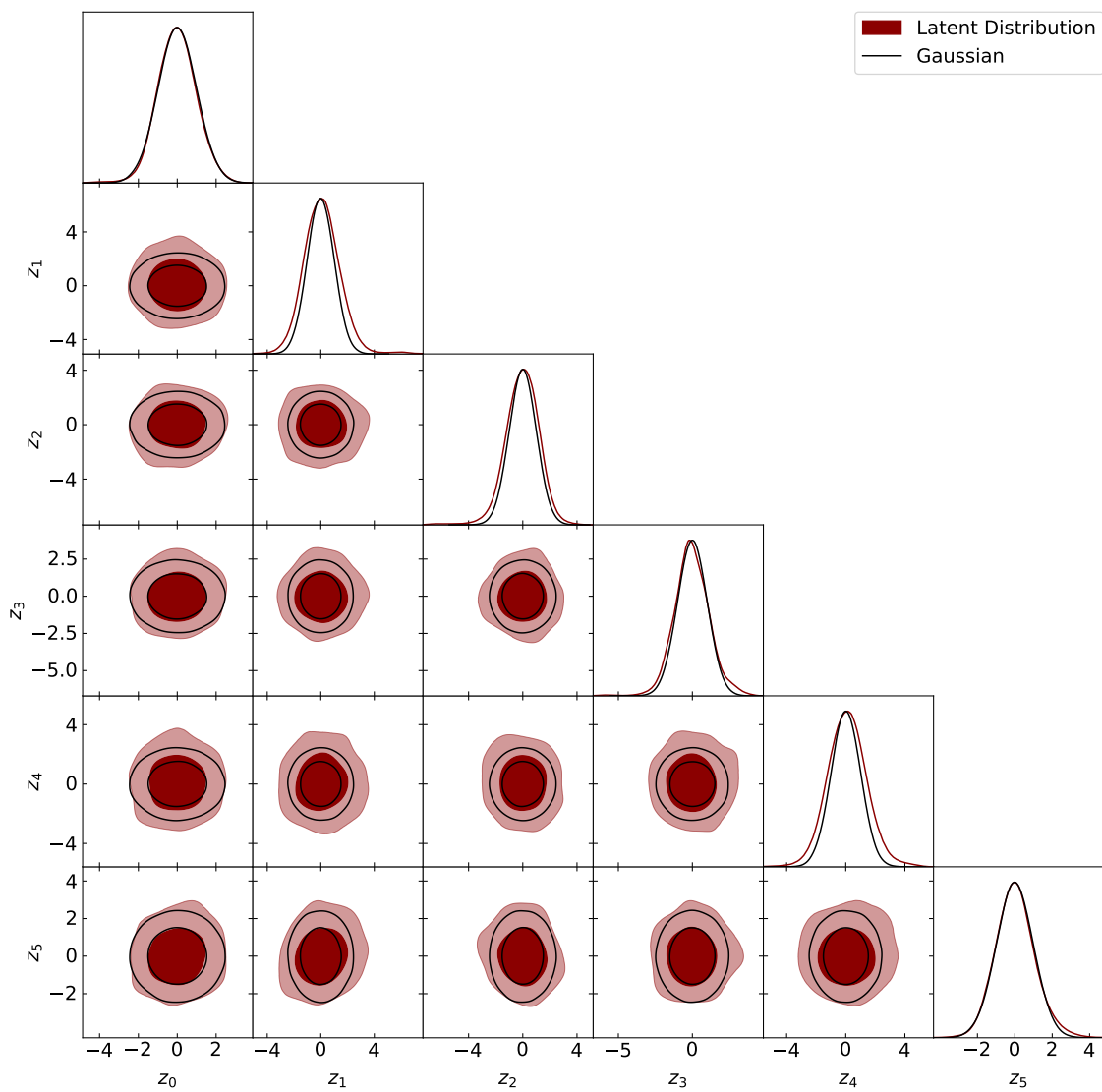


Figure 9.5: Distribution of the latent space variables and a unit Gaussian as comparison.

Finally, one can look at the full posterior, given one simulation, which can be taken as an assumed measurement. The light-cone is chosen semi-randomly from the test data set. Its only requirement is, that its parameters are not at the boundaries of the parameter ranges. The values are

$$\begin{aligned} m_{\text{WDM}} &= 1.71 \text{ keV}, & \Omega_{\text{m}} &= 0.318, & E_0 &= 663 \text{ eV}, \\ L_X &= 10^{40} \text{ erg s}^{-1} M_{\odot} \text{ yr}, & T_{\text{vir}} &= 10^{4.70} \text{ K}, & \zeta &= 51.4, \end{aligned} \tag{9.34}$$

but can be adjusted to anything, within the parameter range one likes. Since the network is already trained, changing the fiducial model and getting the posterior is next to no additional work. The 2D posterior with its correlations is shown in Fig. 9.6 and the 1D in Fig. 9.7. The true value is within the one σ region for all parameters. A comparison to previous work, like Ref. [62] with a full MCMC analysis is difficult, as no noise model is included in this work, so far. However, the 3D CNN from Ref. [52], which is identical to the summary network in this work, performed equally well for light-cones with and without noise. Hence, one might take a look at the contours, but one has to stay cautious. The CL-intervals in this work are of the same order of magnitude as in Ref. [62], showing a successful set-up.

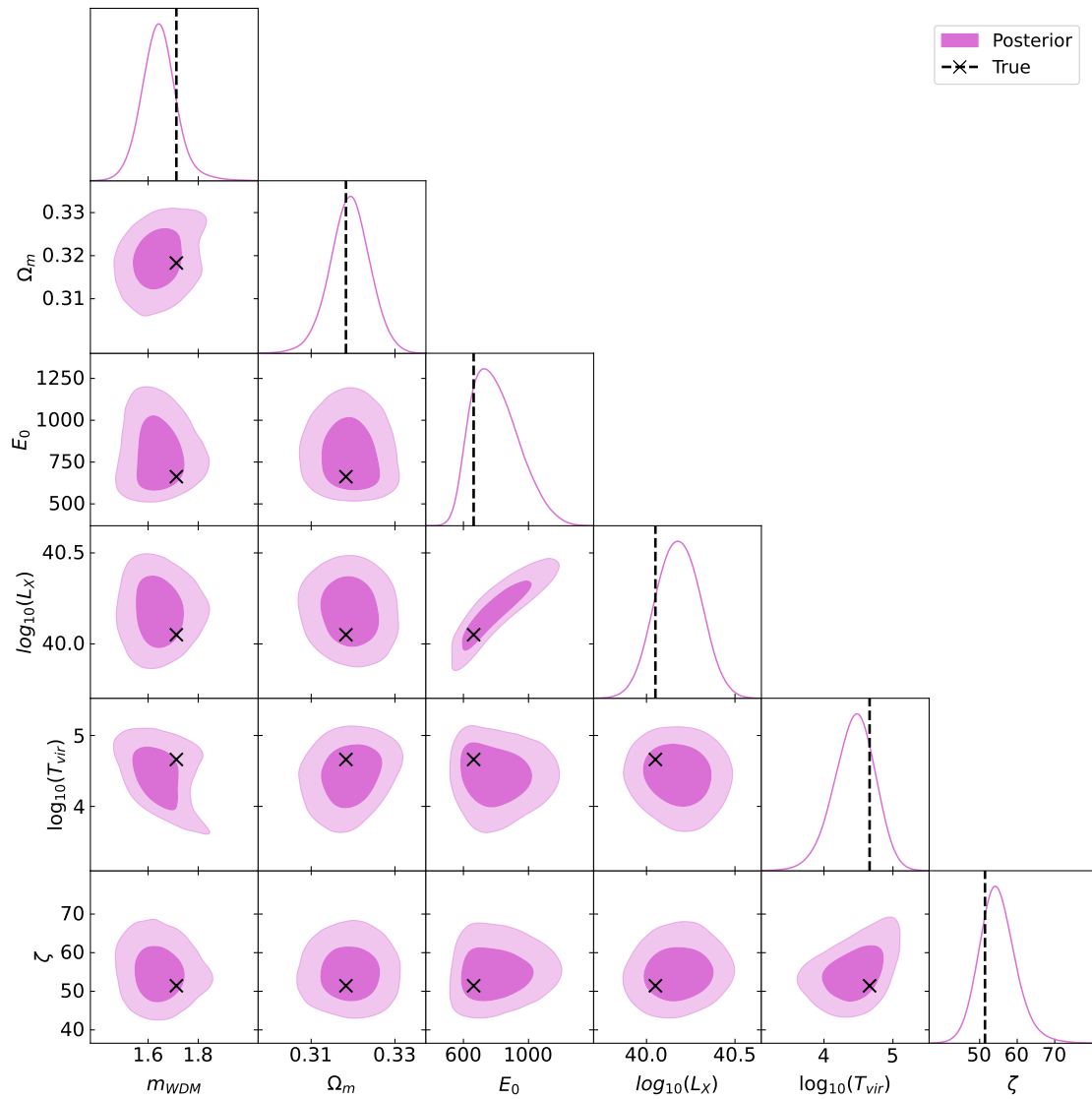


Figure 9.6: Marginalized 2D posterior and true values known from the simulation.

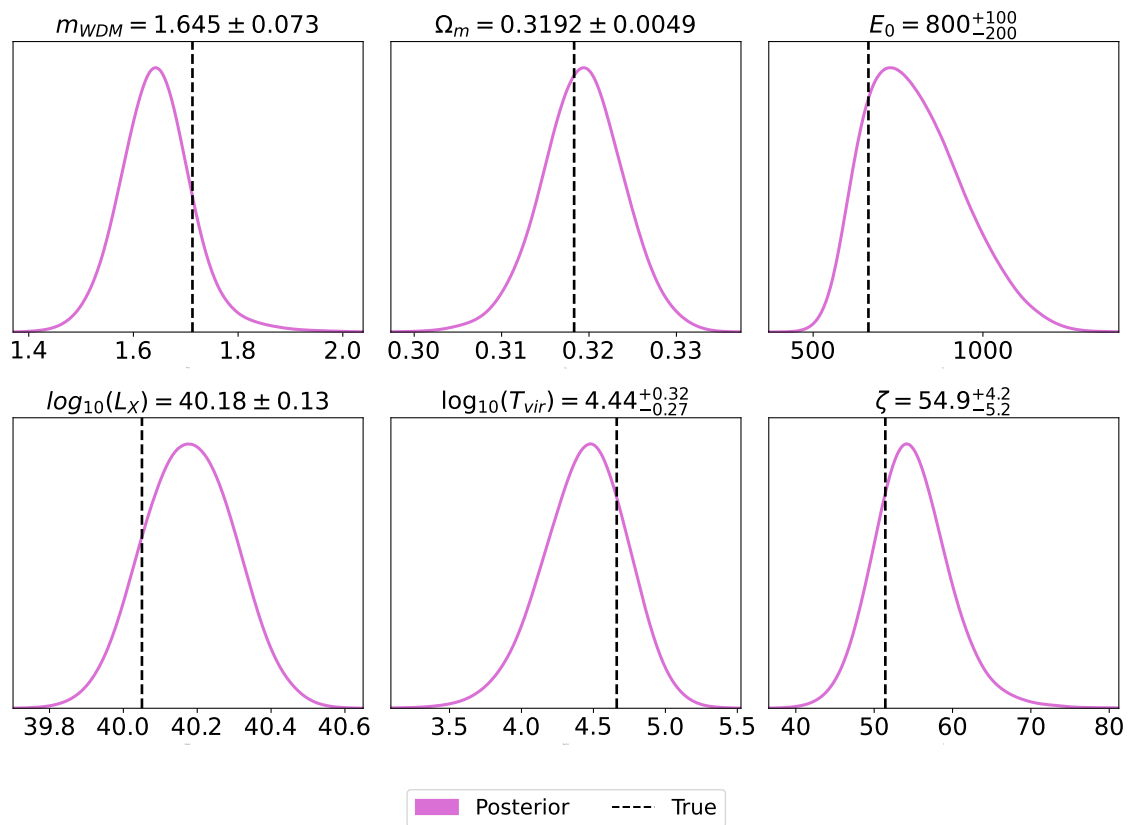


Figure 9.7: Marginalized 1D posterior with 68% CL and true values known from the simulation.

CHAPTER 10

Summary and Conclusion

This thesis starts by discussing the standard model of cosmology and then continues with its problems and how inflation can solve them. The physics of single-field inflation is presented, followed by the theory of cosmological perturbations, which is used to describe the evolution of the quantum mechanical fluctuations during inflation. Before discussing how one can observe the effects of inflation in the CMB and in large scale structure the particular inflationary model of interest is presented, an extension of the Starobinsky model, and the physics discussed. Some analytic motivations are presented to get an intuition for the parameters.

The central task of this thesis is a Markov chain Monte Carlo analysis with multiple likelihoods from future 21cm (SKA) and CMB (LiteBIRD and CMB-S4) experiments. The general algorithm and the tools, MontyPython for the MCMC and CLASS for the evolution of perturbations, are explained, before discussing how the likelihoods are constructed.

First of all the results are consistent with the analysis from the Planck collaboration for the 2018 data. The second Starobinsky parameter c , which describes the extension, is not considered in the Planck analysis, however it needs to be consistent with zero. The results show a slight bias towards non-zero c . The best-fit values from this analysis are used as fiducial values for the forecasts. First, only the combination of the future CMB experiments is considered, already giving a huge improvement in the constraints, excluding $c = 0$ at 95% CL. The different strength of these experiments is visible as well, where the LiteBIRD satellite is superior in constraining τ_{reio} due to its large sky coverage, but for every other parameter the better resolution of the ground base CMB-S4 experiment is more important.

Adding the SKA likelihood improves the results drastically, especially for the cosmological parameters h and ω_{cdm} as well as the second Starobinsky parameter c , excluding a $c = 0$ at 99% CL, if the best-fit value from the Planck analysis is in fact the true value. Therefore, SKA will be able to find hints for physics pointing towards a modification of standard General Relativity.

There exist multiple ideas how to improve the analysis. The likelihood for SKA uses an empirical fitting formula for some of the astrophysical processes, while one could try to implement the full model in MontyPython and CLASS. Of course this comes

with computational cost and challenges as it increases the parameter space, but it would also open new possibilities as it would be possible to consider a larger redshift range, where the fitting formula in this thesis is not applicable. Additionally, the degeneracy between N_* and M might be resolvable with a better description of the reheating process. Again, one could try to implement different reheating scenarios in CLASS

One of those ideas, an improved consideration of astrophysical processes is explored in the end. But more as a proof of concept, where the parameter set changes to EoR and CD parameters with dark matter properties. Now, the inference is done in a likelihood-free way using, directly from the simulated 21cm data. The set-up is self consistent and produces the correct posterior, where the inclusion of DM parameters shows, that fundamental parameters can be inferred with this method. So far, no noise model is included, making it difficult to compare the results to other works who use MCMC. Finally, including a more elaborate inflation mechanism in the simulation will allow the determination of those parameters using likelihood-free inference in the future.

Acknowledgments

First, I would like to thank my supervisors Tilman Plehn and Björn Malte Schäfer for giving me the opportunity to pursue this project and their continued guidance. I am grateful to Lennart Röver and Tanmoy Modak for helping me solve my countless problems along the way. Additionally, I want to thank Caroline Heneka and Theo Heimel for their help in astrophysics and machine learning, respectively. Finally, I want to thank my office mates for all the help and the nice atmosphere.

APPENDIX A

Hubble Slow-Roll Projections

The analysis of this work is very focused on the Starobinsky model of inflation. However, the framework can also be used to constrain inflation in a model independent way, using the Hubble slow-roll (HSR) parameters, defined in Eq. 3.16. The parameter set is then given by

$$\{\omega_b, \omega_{\text{cdm}}, h, \tau_{\text{reio}}, \tilde{A}_s, \epsilon_H, \eta_H, \xi_H^2, \omega_H^3\}. \quad (\text{A.1})$$

The pipeline of the analysis is exactly the same as before. CLASS is used to determine the power spectra from expansion of the Hubble function and MontePython performs the MCMC. First, the constraints from the Planck collaboration are reproduced in Fig. A.1, where only the distribution for the HSR-parameters are shown. Again, the best-fit values are used as fiducial values for the forecasts. For LiteBIRD, CMB-S4 and the combination of Planck with SKA the contours are shown in Fig. A.1, while the combination of LiteBIRD low- ℓ with CMB-S4 high- ℓ as well as LiteBIRD low- ℓ with CMB-S4 high- ℓ and SKA is shown in Fig. A.2. All of the best-fit, mean and 95%CL interval values are given in Tab. A.1.

The qualitative performance of the future experiments is the same as in the model dependent analysis. The best projected constraints come from combination of future 21cm and CMB experiments.

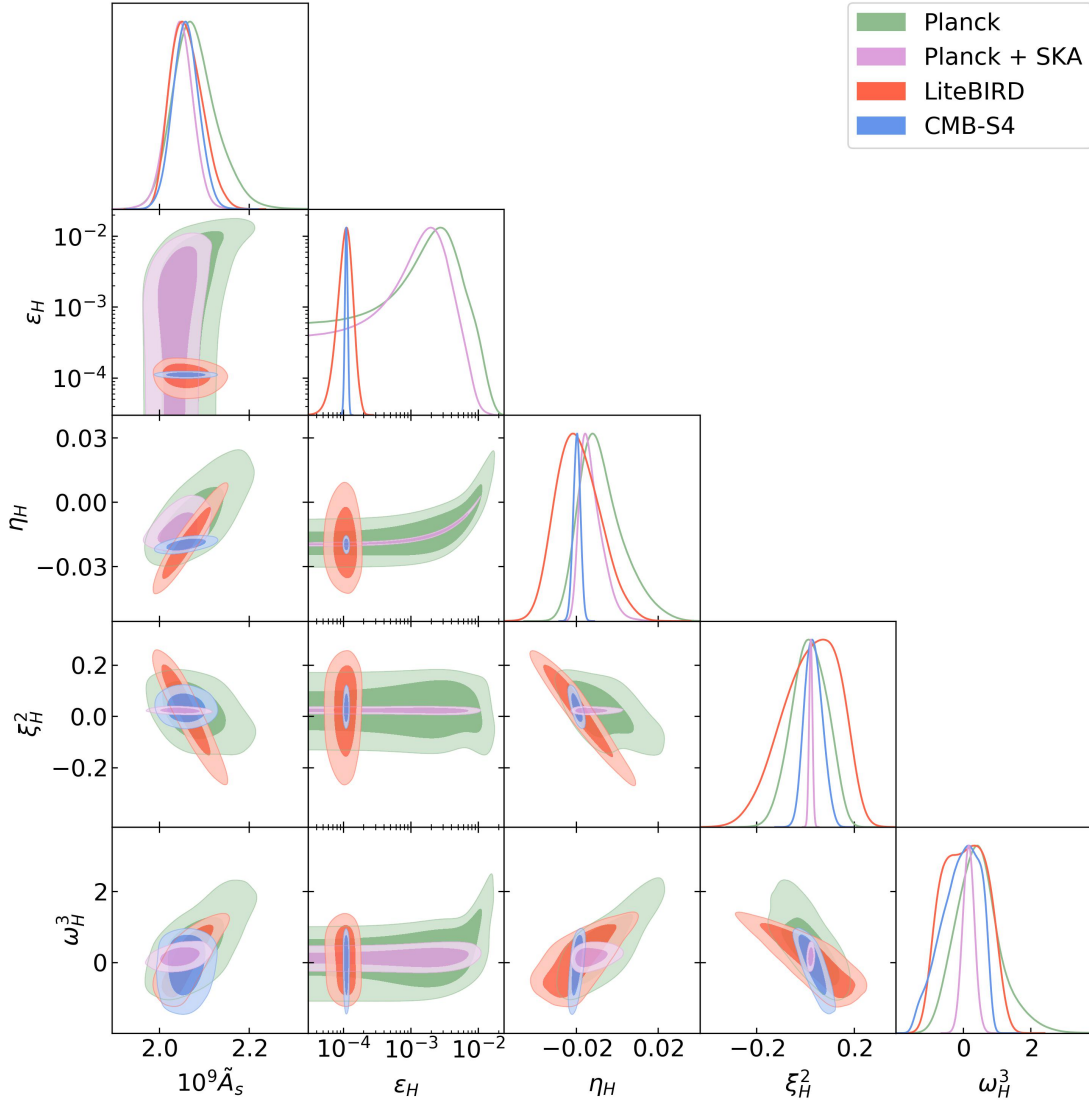


Figure A.1: Marginalized CMB posteriors for the HSR parameters based on Planck (TT , TE , EE +low- ℓ EE +low- ℓ TT), Planck+SKA, LiteBIRD, CMB-S4.

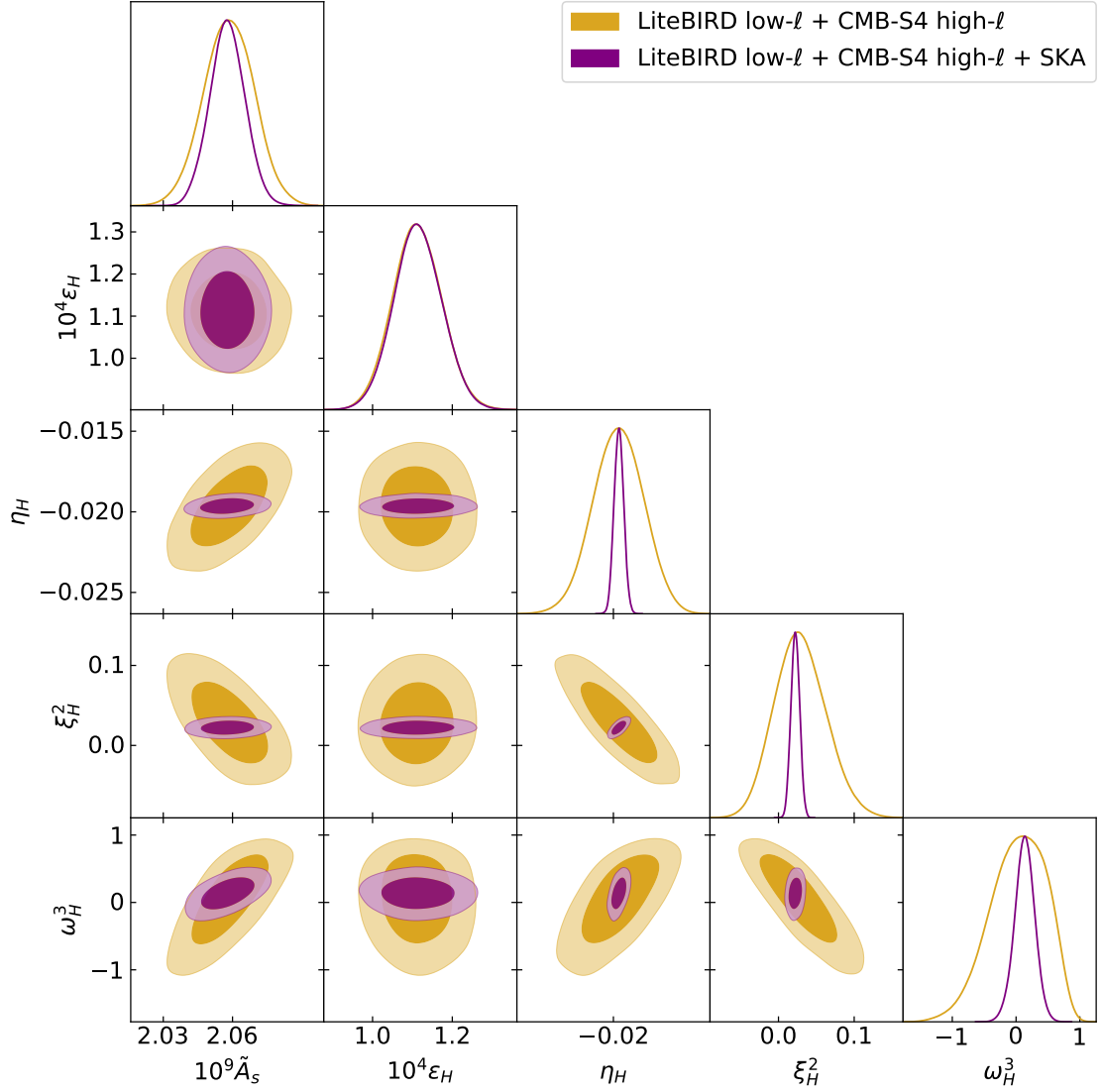


Figure A.2: Marginalized posteriors for the HSR parameters based on LiteBIRD low- ℓ +CMB-S4 high- ℓ and LiteBIRD low- ℓ +CMB-S4 high- ℓ +SKA.), Planck+SKA, LiteBIRD, CMB-S4.

Data	Parameters	Best-fit	Mean $\pm\sigma$	95% lower	95% upper
Planck	$10^9 \tilde{A}_s$	2.059	$2.080^{+0.039}_{-0.057}$	1.987	2.186
	ϵ_H	0.0001111	$0.005445^{+0.002930}_{-0.005313}$	—	< 0.01393
	η_H	-0.01953	$-0.007755^{+0.007628}_{-0.012983}$	-0.02712	0.01599
	ξ_H^2	0.02493	$0.01809^{+0.07614}_{-0.07272}$	-0.1175	0.1587
	ω_H^3	0.1008	$0.4431^{+0.5342}_{-0.7216}$	-0.8662	1.812
Planck + SKA	$10^9 \tilde{A}_s$	2.034	$2.047^{+0.028}_{-0.029}$	1.986	2.102
	ϵ_H	0.001220	$0.003338^{+0.001678}_{-0.003156}$	—	< 0.009013
	η_H	-0.0175	$-0.01283^{+0.00353}_{-0.00658}$	-0.02158	-0.001015
	ξ_H^2	0.02386	$0.02238^{+0.00687}_{-0.00676}$	0.00856	0.0356
	ω_H^3	0.01487	$0.1594^{+0.5060}_{-0.1722}$	-0.1846	0.5023
LiteBIRD	$10^9 \tilde{A}_s$	2.052	$2.061^{+0.030}_{-0.042}$	1.995	2.131
	$10^4 \epsilon_H$	1.086	$1.151^{+0.260}_{-0.316}$	0.5912	1.736
	η_H	-0.01991	$-0.01897^{+0.00938}_{-0.01229}$	-0.03931	0.002793
	ξ_H^2	0.04889	$0.02338^{+0.13927}_{-0.09424}$	-0.2021	0.2296
	ω_H^3	-0.2849	$0.04961^{+0.67578}_{-0.68982}$	-1.059	1.155
CMB-S4	$10^9 \tilde{A}_s$	2.06	$2.059^{+0.028}_{-0.030}$	2.002	2.118
	$10^4 \epsilon_H$	1.104	$1.113^{+0.053}_{-0.057}$	1.007	1.221
	η_H	-0.0191	$-0.0198^{+0.0018}_{-0.0018}$	-0.02334	-0.01632
	ξ_H^2	0.01098	$0.03418^{+0.03331}_{-0.03942}$	-0.03461	0.1062
	ω_H^3	0.3613	$-0.09232^{+0.73948}_{-0.37747}$	-1.157	0.8509
LiteBIRD low- ℓ + CMB-S4 high- ℓ	$10^9 \tilde{A}_s$	2.053	$2.059^{+0.011}_{-0.011}$	2.037	2.08
	$10^4 \epsilon_H$	1.109	$1.112^{+0.060}_{-0.064}$	0.9925	1.235
	η_H	-0.01999	$-0.01966^{+0.00166}_{-0.00165}$	-0.02291	-0.01636
	ξ_H^2	0.02682	$0.0288^{+0.0316}_{-0.0359}$	-0.03628	0.09762
	ω_H^3	-0.003998	$0.03371^{+0.51047}_{-0.37005}$	-0.7958	0.8296
LiteBIRD low- ℓ + CMB-S4 high- ℓ + SKA	$10^9 \tilde{A}_s$	2.06	$2.058^{+0.008}_{-0.008}$	2.042	2.073
	$10^4 \epsilon_H$	1.114	$1.114^{+0.060}_{-0.062}$	0.9967	1.235
	η_H	-0.01951	$-0.01964^{+0.00031}_{-0.00033}$	-0.02026	-0.01902
	ξ_H^2	0.02687	$0.02238^{+0.00568}_{-0.00568}$	0.01128	0.03339
	ω_H^3	0.1544	$0.1347^{+0.1561}_{-0.1525}$	-0.1974	0.4407

Table A.1: Best-fit values, mean, error bars, and 95%CL limits for the HSR parameters shown in Figs. A.1 and A.2.

Bibliography

- [1] Alan H. Guth. “The Inflationary Universe: A Possible Solution to the Horizon and Flatness Problems”. In: *Phys. Rev. D* 23 (1981). Ed. by Li-Zhi Fang and R. Ruffini, pp. 347–356. DOI: [10.1103/PhysRevD.23.347](https://doi.org/10.1103/PhysRevD.23.347).
- [2] K. Sato. “First Order Phase Transition of a Vacuum and Expansion of the Universe”. In: *Mon. Not. Roy. Astron. Soc.* 195 (1981), pp. 467–479.
- [3] Alexei A. Starobinsky. “A New Type of Isotropic Cosmological Models Without Singularity”. In: *Phys. Lett. B* 91 (1980). Ed. by I. M. Khalatnikov and V. P. Mineev, pp. 99–102. DOI: [10.1016/0370-2693\(80\)90670-X](https://doi.org/10.1016/0370-2693(80)90670-X).
- [4] N. Aghanim et al. “Planck 2018 results. VI. Cosmological parameters”. In: *Astron. Astrophys.* 641 (2020). [Erratum: *Astron. Astrophys.* 652, C4 (2021)], A6. DOI: [10.1051/0004-6361/201833910](https://doi.org/10.1051/0004-6361/201833910). arXiv: [1807.06209](https://arxiv.org/abs/1807.06209) [[astro-ph.CO](#)].
- [5] Y. Akrami et al. “Planck 2018 results. X. Constraints on inflation”. In: *Astron. Astrophys.* 641 (2020), A10. DOI: [10.1051/0004-6361/201833887](https://doi.org/10.1051/0004-6361/201833887). arXiv: [1807.06211](https://arxiv.org/abs/1807.06211) [[astro-ph.CO](#)].
- [6] Alexander Vilenkin. “Classical and Quantum Cosmology of the Starobinsky Inflationary Model”. In: *Phys. Rev. D* 32 (1985), p. 2511. DOI: [10.1103/PhysRevD.32.2511](https://doi.org/10.1103/PhysRevD.32.2511).
- [7] Milan B. Mijic, Michael S. Morris, and Wai-Mo Suen. “The R^{*2} Cosmology: Inflation Without a Phase Transition”. In: *Phys. Rev. D* 34 (1986), p. 2934. DOI: [10.1103/PhysRevD.34.2934](https://doi.org/10.1103/PhysRevD.34.2934).
- [8] Kei-ichi Maeda. “Inflation as a Transient Attractor in R^{*2} Cosmology”. In: *Phys. Rev. D* 37 (1988), p. 858. DOI: [10.1103/PhysRevD.37.858](https://doi.org/10.1103/PhysRevD.37.858).
- [9] M. Hazumi et al. “LiteBIRD: A Satellite for the Studies of B-Mode Polarization and Inflation from Cosmic Background Radiation Detection”. In: *J. Low Temp. Phys.* 194.5-6 (2019), pp. 443–452. DOI: [10.1007/s10909-019-02150-5](https://doi.org/10.1007/s10909-019-02150-5).
- [10] T. Matsumura et al. “Mission design of LiteBIRD”. In: *J. Low Temp. Phys.* 176 (2014), p. 733. DOI: [10.1007/s10909-013-0996-1](https://doi.org/10.1007/s10909-013-0996-1). arXiv: [1311.2847](https://arxiv.org/abs/1311.2847) [[astro-ph.IM](#)].
- [11] E. Allys et al. “Probing Cosmic Inflation with the LiteBIRD Cosmic Microwave Background Polarization Survey”. In: (Feb. 2022). DOI: [10.1093/ptep/ptac150](https://doi.org/10.1093/ptep/ptac150). arXiv: [2202.02773](https://arxiv.org/abs/2202.02773) [[astro-ph.IM](#)].

- [12] Kevork Abazajian et al. “CMB-S4: Forecasting Constraints on Primordial Gravitational Waves”. In: *Astrophys. J.* 926.1 (2022), p. 54. DOI: [10.3847/1538-4357/ac1596](https://doi.org/10.3847/1538-4357/ac1596). arXiv: [2008.12619](https://arxiv.org/abs/2008.12619) [[astro-ph.CO](#)].
- [13] Kevork Abazajian et al. “CMB-S4 Science Case, Reference Design, and Project Plan”. In: (July 2019). arXiv: [1907.04473](https://arxiv.org/abs/1907.04473) [[astro-ph.IM](#)].
- [14] Kevork Abazajian et al. “Snowmass 2021 CMB-S4 White Paper”. In: (Mar. 2022). arXiv: [2203.08024](https://arxiv.org/abs/2203.08024) [[astro-ph.CO](#)].
- [15] Kevork N. Abazajian et al. “CMB-S4 Science Book, First Edition”. In: (Oct. 2016). arXiv: [1610.02743](https://arxiv.org/abs/1610.02743) [[astro-ph.CO](#)].
- [16] A. Weltman et al. “Fundamental physics with the Square Kilometre Array”. In: *Publ. Astron. Soc. Austral.* 37 (2020), e002. DOI: [10.1017/pasa.2019.42](https://doi.org/10.1017/pasa.2019.42). arXiv: [1810.02680](https://arxiv.org/abs/1810.02680) [[astro-ph.CO](#)].
- [17] Jonathan Pritchard et al. “Cosmology from EoR/Cosmic Dawn with the SKA”. In: *PoS AASKA14* (2015). Ed. by Tyler L. Bourke et al., p. 012. DOI: [10.22323/1.215.0012](https://doi.org/10.22323/1.215.0012). arXiv: [1501.04291](https://arxiv.org/abs/1501.04291) [[astro-ph.CO](#)].
- [18] David J. Bacon et al. “Cosmology with Phase 1 of the Square Kilometre Array: Red Book 2018: Technical specifications and performance forecasts”. In: *Publ. Astron. Soc. Austral.* 37 (2020), e007. DOI: [10.1017/pasa.2019.51](https://doi.org/10.1017/pasa.2019.51). arXiv: [1811.02743](https://arxiv.org/abs/1811.02743) [[astro-ph.CO](#)].
- [19] Daniel Baumann. “Inflation”. In: *Theoretical Advanced Study Institute in Elementary Particle Physics: Physics of the Large and the Small*. 2011, pp. 523–686. DOI: [10.1142/9789814327183_0010](https://doi.org/10.1142/9789814327183_0010). arXiv: [0907.5424](https://arxiv.org/abs/0907.5424) [[hep-th](#)].
- [20] Andrew R. Liddle, Paul Parsons, and John D. Barrow. “Formalizing the slow roll approximation in inflation”. In: *Phys. Rev. D* 50 (1994), pp. 7222–7232. DOI: [10.1103/PhysRevD.50.7222](https://doi.org/10.1103/PhysRevD.50.7222). arXiv: [astro-ph/9408015](https://arxiv.org/abs/astro-ph/9408015).
- [21] V.F. Mukhanov, H.A. Feldman, and R.H. Brandenberger. “Theory of cosmological perturbations”. In: *Physics Reports* 215.5 (1992), pp. 203–333. ISSN: 0370-1573. DOI: [https://doi.org/10.1016/0370-1573\(92\)90044-Z](https://doi.org/10.1016/0370-1573(92)90044-Z). URL: <https://www.sciencedirect.com/science/article/pii/037015739290044Z>.
- [22] Bruce A. Bassett, Shinji Tsujikawa, and David Wands. “Inflation dynamics and reheating”. In: *Rev. Mod. Phys.* 78 (2006), pp. 537–589. DOI: [10.1103/RevModPhys.78.537](https://doi.org/10.1103/RevModPhys.78.537). arXiv: [astro-ph/0507632](https://arxiv.org/abs/astro-ph/0507632).
- [23] Salvatore Capozziello and Mariafelicia De Laurentis. “Extended Theories of Gravity”. In: *Phys. Rept.* 509 (2011), pp. 167–321. DOI: [10.1016/j.physrep.2011.09.003](https://doi.org/10.1016/j.physrep.2011.09.003). arXiv: [1108.6266](https://arxiv.org/abs/1108.6266) [[gr-qc](#)].
- [24] Vsevolod R. Ivanov et al. “Analytic extensions of Starobinsky model of inflation”. In: *JCAP* 03.03 (2022), p. 058. DOI: [10.1088/1475-7516/2022/03/058](https://doi.org/10.1088/1475-7516/2022/03/058). arXiv: [2111.09058](https://arxiv.org/abs/2111.09058) [[gr-qc](#)].
- [25] Antonio De Felice and Shinji Tsujikawa. “f(R) theories”. In: *Living Rev. Rel.* 13 (2010), p. 3. DOI: [10.12942/lrr-2010-3](https://doi.org/10.12942/lrr-2010-3). arXiv: [1002.4928](https://arxiv.org/abs/1002.4928) [[gr-qc](#)].
- [26] Scott Dodelson. *Modern Cosmology*. 2003.
- [27] Tim Sprenger et al. “Cosmology in the era of Euclid and the Square Kilometre Array”. In: *JCAP* 02 (2019), p. 047. DOI: [10.1088/1475-7516/2019/02/047](https://doi.org/10.1088/1475-7516/2019/02/047). arXiv: [1801.08331](https://arxiv.org/abs/1801.08331) [[astro-ph.CO](#)].

- [28] Yi Mao et al. “How accurately can 21 cm tomography constrain cosmology?” In: *Phys. Rev. D* 78 (2008), p. 023529. DOI: [10.1103/PhysRevD.78.023529](https://doi.org/10.1103/PhysRevD.78.023529). arXiv: [0802.1710](https://arxiv.org/abs/0802.1710) [[astro-ph](#)].
- [29] Raul A. Monsalve. “Synchrotron Radiation as a Foreground to the Global Redshifted 21-cm Measurement by EDGES”. Radio Synchrotron Background Conference. 2017. URL: <https://physics.richmond.edu/rsb-conference/presentation-pdfs/RSB%20-%20Monsalve.pdf>.
- [30] Matias Zaldarriaga, Steven R. Furlanetto, and Lars Hernquist. “21 Centimeter fluctuations from cosmic gas at high redshifts”. In: *Astrophys. J.* 608 (2004), pp. 622–635. DOI: [10.1086/386327](https://doi.org/10.1086/386327). arXiv: [astro-ph/0311514](https://arxiv.org/abs/astro-ph/0311514).
- [31] Julian B. Muñoz, Yacine Ali-Haïmoud, and Marc Kamionkowski. “Primordial non-gaussianity from the bispectrum of 21-cm fluctuations in the dark ages”. In: *Phys. Rev. D* 92.8 (2015), p. 083508. DOI: [10.1103/PhysRevD.92.083508](https://doi.org/10.1103/PhysRevD.92.083508). arXiv: [1506.04152](https://arxiv.org/abs/1506.04152) [[astro-ph.CO](#)].
- [32] Tanmoy Modak et al. “Cornering Extended Starobinsky Inflation with CMB and SKA”. In: (Oct. 2022). arXiv: [2210.05698](https://arxiv.org/abs/2210.05698) [[astro-ph.CO](#)].
- [33] Pablo Lemos, Anthony Challinor, and George Efstathiou. “The effect of Limber and flat-sky approximations on galaxy weak lensing”. In: *JCAP* 05 (2017), p. 014. DOI: [10.1088/1475-7516/2017/05/014](https://doi.org/10.1088/1475-7516/2017/05/014). arXiv: [1704.01054](https://arxiv.org/abs/1704.01054) [[astro-ph.CO](#)].
- [34] J. C. Jackson. “Fingers of God: A critique of Rees’ theory of primordial gravitational radiation”. In: *Mon. Not. Roy. Astron. Soc.* 156 (1972), 1P–5P. DOI: [10.1093/mnras/156.1.1P](https://doi.org/10.1093/mnras/156.1.1P). arXiv: [0810.3908](https://arxiv.org/abs/0810.3908) [[astro-ph](#)].
- [35] A. Gelman et al. *Bayesian Data Analysis, Third Edition*. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis, 2013. ISBN: 9781439840955. URL: <https://books.google.de/books?id=ZXL6AQAAQBAJ>.
- [36] W. K. Hastings. “Monte Carlo sampling methods using Markov chains and their applications”. In: *Biometrika* 57.1 (Apr. 1970), pp. 97–109. ISSN: 0006-3444. DOI: [10.1093/biomet/57.1.97](https://doi.org/10.1093/biomet/57.1.97). eprint: <https://academic.oup.com/biomet/article-pdf/57/1/97/23940249/57-1-97.pdf>. URL: <https://doi.org/10.1093/biomet/57.1.97>.
- [37] Andrew Gelman and Donald B. Rubin. “Inference from Iterative Simulation Using Multiple Sequences”. In: *Statist. Sci.* 7 (1992), pp. 457–472. DOI: [10.1214/ss/1177011136](https://doi.org/10.1214/ss/1177011136).
- [38] Thejs Brinckmann and Julien Lesgourgues. “MontePython 3: boosted MCMC sampler and other features”. In: *Phys. Dark Univ.* 24 (2019), p. 100260. DOI: [10.1016/j.dark.2018.100260](https://doi.org/10.1016/j.dark.2018.100260). arXiv: [1804.07261](https://arxiv.org/abs/1804.07261) [[astro-ph.CO](#)].
- [39] Joanna Dunkley et al. “Fast and reliable mcmc for cosmological parameter estimation”. In: *Mon. Not. Roy. Astron. Soc.* 356 (2005), pp. 925–936. DOI: [10.1111/j.1365-2966.2004.08464.x](https://doi.org/10.1111/j.1365-2966.2004.08464.x). arXiv: [astro-ph/0405462](https://arxiv.org/abs/astro-ph/0405462).
- [40] Laurence Perotto et al. “Probing cosmological parameters with the CMB: Forecasts from full Monte Carlo simulations”. In: *JCAP* 10 (2006), p. 013. DOI: [10.1088/1475-7516/2006/10/013](https://doi.org/10.1088/1475-7516/2006/10/013). arXiv: [astro-ph/0606227](https://arxiv.org/abs/astro-ph/0606227).
- [41] Thejs Brinckmann et al. “The promising future of a robust cosmological neutrino mass measurement”. In: *JCAP* 01 (2019), p. 059. DOI: [10.1088/1475-7516/2019/01/059](https://doi.org/10.1088/1475-7516/2019/01/059). arXiv: [1808.05955](https://arxiv.org/abs/1808.05955) [[astro-ph.CO](#)].

- [42] Max Tegmark and Matias Zaldarriaga. “The Fast Fourier Transform Telescope”. In: *Phys. Rev. D* 79 (2009), p. 083530. DOI: [10.1103/PhysRevD.79.083530](https://doi.org/10.1103/PhysRevD.79.083530). arXiv: [0805.4414](https://arxiv.org/abs/0805.4414) [astro-ph].
- [43] Julien Lesgourgues. *The Cosmic Linear Anisotropy Solving System (CLASS) III: Comparison with CAMB for LambdaCDM*. 2011. DOI: [10.48550/ARXIV.1104.2934](https://doi.org/10.48550/ARXIV.1104.2934). URL: <https://arxiv.org/abs/1104.2934>.
- [44] Antony Lewis. *GetDist: a Python package for analysing Monte Carlo samples*. 2019. DOI: [10.48550/ARXIV.1910.13970](https://doi.org/10.48550/ARXIV.1910.13970). URL: <https://arxiv.org/abs/1910.13970>.
- [45] Dhong Yeon Cheong, Hyun Min Lee, and Seong Chan Park. “Beyond the Starobinsky model for inflation”. In: *Phys. Lett. B* 805 (2020), p. 135453. DOI: [10.1016/j.physletb.2020.135453](https://doi.org/10.1016/j.physletb.2020.135453). arXiv: [2002.07981](https://arxiv.org/abs/2002.07981) [hep-ph].
- [46] G. Rodrigues-da-Silva, J. Bezerra-Sobrinho, and L. G. Medeiros. “Higher-order extension of Starobinsky inflation: Initial conditions, slow-roll regime, and reheating phase”. In: *Phys. Rev. D* 105.6 (2022), p. 063504. DOI: [10.1103/PhysRevD.105.063504](https://doi.org/10.1103/PhysRevD.105.063504). arXiv: [2110.15502](https://arxiv.org/abs/2110.15502) [astro-ph.CO].
- [47] Alexey S. Koshelev, K. Sravan Kumar, and Alexei A. Starobinsky. “Generalized non-local R^2 -like inflation”. In: (Sept. 2022). arXiv: [2209.02515](https://arxiv.org/abs/2209.02515) [hep-th].
- [48] Tanmoy Modak et al. “Probing the Inflaton Potential with SKA”. In: *SciPost Phys. Core* 5 (2022), p. 037. DOI: [10.21468/SciPostPhysCore.5.3.037](https://doi.org/10.21468/SciPostPhysCore.5.3.037). arXiv: [2112.09148](https://arxiv.org/abs/2112.09148) [astro-ph.CO].
- [49] Steven G. Murray et al. “21cmFAST v3: A Python-integrated C code for generating 3D realizations of the cosmic 21cm signal.” In: *Journal of Open Source Software* 5.54 (2020), p. 2582. DOI: [10.21105/joss.02582](https://doi.org/10.21105/joss.02582). URL: <https://doi.org/10.21105/joss.02582>.
- [50] Andrei Mesinger, Steven Furlanetto, and Renyue Cen. “21cmfast: a fast, seminumerical simulation of the high-redshift 21-cm signal”. In: *Monthly Notices of the Royal Astronomical Society* 411.2 (Feb. 2011), pp. 955–972. ISSN: 0035-8711. DOI: [10.1111/j.1365-2966.2010.17731.x](https://doi.org/10.1111/j.1365-2966.2010.17731.x). eprint: <https://academic.oup.com/mnras/article-pdf/411/2/955/4099991/mnras0411-0955.pdf>. URL: <https://doi.org/10.1111/j.1365-2966.2010.17731.x>.
- [51] Ya. B. Zel’dovich. “Gravitational instability: An approximate theory for large density perturbations.” In: 5 (Mar. 1970), pp. 84–89. URL: <https://ui.adsabs.harvard.edu/abs/1970A&A....5...84Z>.
- [52] Steffen Neutsch, Caroline Heneka, and Marcus Brüggen. “Inferring astrophysics and dark matter properties from 21 cm tomography using deep learning”. In: *Mon. Not. Roy. Astron. Soc.* 511.3 (2022), pp. 3446–3462. DOI: [10.1093/mnras/stac218](https://doi.org/10.1093/mnras/stac218). arXiv: [2201.07587](https://arxiv.org/abs/2201.07587) [astro-ph.CO].
- [53] Tilman Plehn. *Modern Machine Learning for LHC Physicists*. Feb. 2023. URL: https://www.thphys.uni-heidelberg.de/~plehn/pics/modern_ml.pdf.
- [54] Diederik P. Kingma and Jimmy Ba. *Adam: A Method for Stochastic Optimization*. 2017. arXiv: [1412.6980](https://arxiv.org/abs/1412.6980) [cs.LG].

- [55] Stefan T. Radev et al. *BayesFlow: Learning complex stochastic models with invertible neural networks*. 2020. DOI: [10.48550/ARXIV.2003.06281](https://doi.org/10.48550/ARXIV.2003.06281). URL: <https://arxiv.org/abs/2003.06281>.
- [56] Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. *Density estimation using Real NVP*. 2017. arXiv: [1605.08803](https://arxiv.org/abs/1605.08803) [cs.LG].
- [57] Lynton Ardizzone et al. *Framework for Easily Invertible Architectures (FrEIA)*. 2018-2022. URL: <https://github.com/vislearn/FrEIA>.
- [58] Adam Paszke et al. “Automatic differentiation in PyTorch”. In: (2017).
- [59] Sean Talts et al. *Validating Bayesian Inference Algorithms with Simulation-Based Calibration*. 2018. DOI: [10.48550/ARXIV.1804.06788](https://doi.org/10.48550/ARXIV.1804.06788). URL: <https://arxiv.org/abs/1804.06788>.
- [60] Lynton Ardizzone et al. *Analyzing Inverse Problems with Invertible Neural Networks*. 2018. DOI: [10.48550/ARXIV.1808.04730](https://doi.org/10.48550/ARXIV.1808.04730). URL: <https://arxiv.org/abs/1808.04730>.
- [61] Anchal Saxena et al. “Impact of dark matter models on the EoR 21-cm signal bispectrum”. English. In: *Monthly Notices of the Royal Astronomical Society* 497.3 (June 2020), pp. 2941–2953. ISSN: 0035-8711. DOI: [10.1093/mnras/staa1768](https://doi.org/10.1093/mnras/staa1768).
- [62] Bradley Greig and Andrei Mesinger. “21CMMC with a 3D light-cone: the impact of the co-evolution approximation on the astrophysics of reionization and cosmic dawn”. In: *Monthly Notices of the Royal Astronomical Society* 477.3 (Mar. 2018), pp. 3217–3229. ISSN: 0035-8711. DOI: [10.1093/mnras/sty796](https://doi.org/10.1093/mnras/sty796). eprint: <https://academic.oup.com/mnras/article-pdf/477/3/3217/24802887/sty796.pdf>. URL: <https://doi.org/10.1093/mnras/sty796>.

Erklärung:

Ich versichere, dass ich diese Arbeit selbstständig verfasst habe und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Heidelberg, den 03.04.2023

B. Jha
.....