



THEORETISCHE PHYSIK

Gauge-Field Production in Axion Inflation

—

From Background Dynamics to Primordial Perturbations
with the Gradient-Expansion Formalism

INAUGURAL-DISSERTATION
ZUR ERLANGUNG DES DOKTORGRADES (DR. RER. NAT.)
DER NATURWISSENSCHAFTEN IM FACHBEREICH DER PHYSIK
DER MATHEMATISCH-NATURWISSENSCHAFTLICHEN FAKULTÄT
DER UNIVERSITÄT MÜNSTER

vorgelegt von
Richard von Eckardstein
aus Münster
— 2026 —

Abstract

Cosmic inflation is the leading paradigm for explaining the origin of the primordial perturbations that seed the anisotropies observed in the cosmic microwave background and later grow into the large-scale structure of the Universe. In the conventional model of slow-roll inflation, these primordial perturbations, together with primordial gravitational waves (GW), are generated from vacuum fluctuations of the spacetime metric and the scalar inflaton field that drives inflation. An appealing extension of this framework is axion inflation. In this model, the inflaton field is a pseudoscalar coupled to an Abelian gauge field via the shift-symmetric Chern–Simons operator $\phi F\tilde{F}$. Through this interaction, the rolling inflaton excites and amplifies helical gauge fields during inflation, which in turn directly source enhanced primordial perturbations. Among other effects, this can amplify inflationary GWs to a level where they may be observable with current or next-generation GW experiments.

However, making quantitatively accurate predictions in axion inflation is challenging, since the underlying dynamical equations are highly nonlinear. Through their interaction, the inflaton and the gauge field continuously exchange energy, which not only drives gauge-field production but also induces backreaction on the inflaton. Moreover, in realistic extensions of this model, the gauge field may itself couple to charged fields. This can lead to the direct production of charged particles during inflation via the Schwinger effect. While this is phenomenologically appealing, for instance as a mechanism for dark matter production during inflation, it also increases the dynamical complexity of the model, since the produced charged particles can themselves backreact on gauge-field production. Such effects must be self-consistently accounted for to determine relevant observable signatures, such as the spectrum of primordial GWs from axion inflation.

This thesis presents four research projects connected by the common requirement of accurately modelling gauge-field production in axion inflation. The backbone of these analyses is the *gradient-expansion formalism* (GEF), a semi-analytical method designed to describe the nontrivial interplay between the inflaton field, the gauge field, and possible charged particles.

The first two analyses employ the GEF to study the backreaction of gauge-field production on the inflaton field. The first revisits the stability of the Anber–Sorbo solution, which describes a stable attractor in which gauge-field production is perfectly stabilized and the inflaton follows a steady-state evolution sustained entirely by gauge-field friction. We substantiate existing claims that this solution is unstable, but we also identify previously unexplored regions of axion-inflation parameter space in which the Anber–Sorbo solution may be stable.

The second study revisits the effective treatment of the Schwinger effect in axion inflation, in which the backreaction of charged particles is modelled as a conductive medium that dampens gauge-field production. We propose a refined effective description that accounts for two effects. First, we relax an idealizing assumption that Schwinger pair production in axion inflation proceeds through perfectly collinear electric and magnetic fields. Second, we implement Schwinger damping as a scale-dependent effect, such that only gauge-field modes with wavelengths above the typical separation of charged particles in the conductive medium are subject to damping. In a comparative analysis with existing approaches in the literature, we demonstrate that these modifications can lead to order-of-magnitude differences in the energy density of gauge fields and charged particles in axion inflation. In addition, we identify a new dynamical regime in which strong Schwinger damping and gauge-field backreaction coexist, whereas earlier results had suggested that gauge-field backreaction is fully suppressed once Schwinger damping becomes important.

The final two studies use the GEF to determine the complex inflationary background and, on that basis, compute the effect of gauge-field production on primordial perturbations. The first of these revisits the impact of gauge fields on the generation of primordial curvature perturbations. A formalism is developed that consistently includes effects previously neglected in this context,

namely the direct sourcing of metric perturbations by gauge fields and the backreaction of gauge fields on the perturbations themselves. In a first application of this formalism, we show that gauge-field-induced metric perturbations remain subdominant in the total spectrum of curvature perturbations as long as backreaction on the perturbations is neglected.

Finally, we employ the GEF to study the production of primordial GWs. Building on our revised effective treatment of the Schwinger effect, we compare two scenarios of axion inflation: one in which the Schwinger effect is neglected and one in which it effectively produces Standard Model fermions during inflation. For both models of axion inflation, we assess the detectability of gauge-field-induced GWs in current or next-generation GW observatories through a systematic parameter-space analysis. We find qualitatively different behavior in the two scenarios. Without Schwinger damping, gauge-field production is either too weak to yield observable GWs or becomes so efficient that it drives a dynamical instability and ultimately overproduces GWs in conflict with existing cosmological bounds on their amplitude. By contrast, when Schwinger pair production of Standard Model fermions is included, the resulting damping can prevent the onset of this instability, thereby opening regions of parameter space in which GWs from axion inflation may be observable with future ground-based or space-based GW interferometers.

Collectively, the four analyses presented in this thesis develop the GEF into a practical framework for studying aspects of axion inflation whose phenomenology depends sensitively on nonlinear gauge-field dynamics. In this way, the thesis connects the treatment of background evolution, Schwinger pair production, primordial GW production, and scalar perturbations within a common semi-analytical approach, thereby opening the possibility for more detailed analyses of the rich phenomenology of axion inflation in future work.

Acknowledgements

I would like to extend my sincerest gratitude to all the people who supported me throughout the years I spent working towards my PhD.

My first thanks go to my advisor Kai Schmitz for giving me the opportunity to pursue my PhD in his research group. I was fortunate to have been among its first members, and to have experienced firsthand how Kai's dedication, enthusiasm, and commitment to both his work and his students shaped the group into what it is today. Throughout this time, Kai provided me with guidance and support, while also encouraging me to always aim higher. For, this I am deeply grateful.

I would also like to thank Oleksandr Sobol, who was in many ways a second advisor to my PhD. All projects presented in this thesis were developed in close collaboration with him, and I benefited enormously from his insight and experience.

My sincere thanks also extend to Ruth Durrer, Deepen Garg, Marco Peloso, Lorenzo Sorbo, and Stanislav Vilchinskii for our fruitful collaboration on two of the articles I present in this thesis.

During my PhD, I was lucky to have had the opportunity to pursue an extended research stay at West Virginia University, and I want to thank Maura McLaughlin for hosting me. During this stay, I had the great pleasure of meeting many delightful people, and I want to especially thank Antik Sihi and Jackson Taylor for making my time in West Virginia as enjoyable as it was.

I also would like to thank Marco Peloso for hosting me at the University of Padova for a brief research visit there.

Pursuing a PhD can be a dull affair without good colleagues to accompany it, and I was fortunate enough to have had many of them. For our shared time together, I want to thank Amin Abouibrahim, Adeela Afzal, Olivia Bitcon, Jonah Fabian, Adrian Finke, Doa Hashemi, Pia Leonie Jones, Justus Kuhlmann, Laura Moreno Valero, Johannes Pirsch, Anne Poppe, Peter Risse, Rafael Lino dos Santos, Colum de Vries, Luca Paolo Wiggering and Jan Wissmann.

My special gratitude is also to Jan-Christopher Knetsch and Katrin Greve who consistently stood by my side as both colleagues and friends while I was writing this thesis.

Most of all, I want to thank Tobias Schröder. Not only have we been colleagues and office mates for three years, but through our shared experience we have grown into close friends. I will fondly remember the many discussions and laughs that we shared throughout this time.

At this point, I would also like to thank my colleagues who have taken time out of their day to proofread this thesis and to provide me with valuable feedback. My gratitude goes to Jan Christopher Knetsch, Justus Kuhlmann, Paolo Pianizzola, Oleksandr Sobol, Tobias Schröder and Colum de Vries.

Finally, I am lucky to have many people dear to me who always stand by my side. I want to thank Flippi, Clu, Eros, Ella, Silvia, Reto, Pauli, and Rino for all their love and support during my PhD. I also want to thank my family for providing me the stability that I needed in my academic pursuit. My love goes to my father Arnold, my mother Sigrid, my sister Lorenza, and my brother Leonard. And never forget the dogs, Lara and Yuri, for what would life be without them?

List of Publications

The following lists the author's contributions to the original publications/preprints that are included in this thesis. The articles are given in chronological order:

- [1] R. von Eckardstein, M. Peloso, K. Schmitz, O. Sobol, L. Sorbo; *Axion inflation in the strong-backreaction regime: decay of the Anber-Sorbo solution*, *JHEP* 11 (2023) 183 [2309.04254]

This work was conceptualized by K. Schmitz and O. Sobol. I actively participated in these discussions. Together, O. Sobol and I developed the strategy to initialize the gradient-expansion formalism (GEF) in the Anber–Sorbo solution. The results and figures in this article were produced by O. Sobol, while I partially cross-checked them and produced the data for Figs. 9 and 10. The manuscript was written by O. Sobol, K. Schmitz and M. Peloso, while L. Sorbo and I helped finalize it.

- [2] R. Durrer, R. von Eckardstein, D. Garg, K. Schmitz, O. Sobol, S. Vilchinskii; *Scalar perturbations from inflation in the presence of gauge fields*, *Phys. Rev. D* 110 (2024) 4, 043533 [2404.19694]

This work was initiated by R. Durrer, O. Sobol and S. Vilchinskii. The formalism present in this article was devised by O. Sobol and S. Vilchinskii while D. Garg and I realized its numerical implementation. I produced the final numerical data for this article which D. Garg cross-checked. All figures in this article were created by myself. The manuscript was written by R. Durrer, O. Sobol, D. Garg, and S. Vilchinskii, while K. Schmitz and I helped finalize it.

- [3] R. von Eckardstein, K. Schmitz, O. Sobol; *On the Schwinger effect during axion inflation*, *JHEP* 02 (2025) 096 [2408.16538]

This work was initialized by K. Schmitz and O. Sobol. The idea for the mixed picture representation of the induced Schwinger current was brought forth by O. Sobol and K. Schmitz. O. Sobol and I independently derived the corresponding expressions. The idea for the Schwinger pair-creation scale was proposed by O. Sobol. I devised its implementation into the equations of motion for axion inflation. The implementation of the revised effective model into the GEF was worked out by myself with input from O. Sobol. All results in this article were produced by myself, while O. Sobol partially cross-checked them. I created all figures in this article. O. Sobol and I wrote the manuscript and finalized it together with K. Schmitz.

- [4] R. von Eckardstein, K. Schmitz, O. Sobol; *Gravitational waves from axion inflation in the gradient expansion formalism. Part I. Pure axion inflation*, *JHEP* 01 (2026) 018 [2508.00798]

This work was initialized by K. Schmitz, O. Sobol and myself. I derived all expressions and formulae in the article, realized their numerical implementation, produced all results and created all figures. The manuscript was written by myself and finalized with the aid of K. Schmitz, O. Sobol.

- [5] R. von Eckardstein, K. Schmitz, O. Sobol; *Gravitational waves from axion inflation in the gradient expansion formalism. Part II. Fermionic axion inflation*, *JHEP* 03 (2026) 072 [2509.25013]

This work was initialized by K. Schmitz, O. Sobol and myself. I derived all expressions and formulae in the article, realized their numerical implementation, produced all results and created all figures. The manuscript was written by myself and finalized with the aid of K. Schmitz, O. Sobol.

- [6] O. Sobol, R. von Eckardstein, E. Koch, S. Gurevich, U. Thiele, K. Schmitz; *Nonlinear physics of axion inflation* [2603.02570] (preprint)

This work was initialized by K. Schmitz, O. Sobol as a continuation of Ref. [1]. All results and figures in this article were produced by O. Sobol. I participated in discussions that lead to its conceptualization, and my software package [7] was used to produce the results in Fig. 6 in this article. The manuscript was written and finalized by O. Sobol, while I added minor comments.

The following software package was developed alongside this thesis:

- [7] R. von Eckardstein; *GEFF: The Gradient Expansion Formalism Factory* [2510.12644]
 GitHub repository: github.com/richard-von-eckardstein/GEFF
 Documentation: richard-von-eckardstein.github.io/GEFF

This software package was fully developed by myself. I wrote the documentation and the manual. The package was used to obtain the results in Refs. [4, 5] and partially those of Ref. [6]. It was also used to obtain several results and figures in this thesis which were not included in the original publications listed above.

Furthermore, the author contributed to the following publications during his doctoral training, but they are not included in this thesis:

- [8] G. Weymann-Despres, S. Henrot-Versillé, G. Moulta, V. Vennin, L. Duflot, R. von Eckardstein; *MSSM inflation revisited: Toward a coherent description of high-energy physics and cosmology*, *Phys. Rev. D* 108 (2023) 2, 023511 [2304.04534]
 [9] NANOGrav Collaboration (A. Afzal et al.); *The NANOGrav 15 yr Data Set: Search for Signals from New Physics*, *Astrophys. J. Lett.* 951 (2023) 1, L11, *Astrophys. J. Lett.* 971 (2024) 1, L27 (erratum) [2306.16219]
 [10] NANOGrav Collaboration (G. Agazie et al.); *The NANOGrav 15 yr Data Set: Running of the Spectral Index*, *Astrophys. J. Lett.* 978 (2025) 2, L29 01 (2023) 11 [2408.10166]

The author also contributed to the development of the following software package for pulsar-timing-array data analysis which was used in Refs. [9, 10]:

- [11] A. Mitridate, D. Wright, R. von Eckardstein, T. Schröder, J. Nay, K. Olum, K. Schmitz, T. Trickle; *PTArcade* [2306.16377]
 GitHub repository: github.com/andrea-mitridate/PTArcade
 Documentation: andrea-mitridate.github.io/PTArcade/

Conventions

The following conventions are consistently maintained throughout this thesis.

In our metric signature, the Minkowski metric is given by $\eta_{\mu\nu} = \text{diag}(+, -, -, -)$. We will furthermore use Greek indices for four-vectors and Latin indices for Euclidean three-vectors. Additionally, we define the Levi-Civita symbol in four dimensions such that $\varepsilon^{0123} = 1$.

Our convention for the flat Friedmann–Lemaître–Robertson–Walker metric is to write the line element as

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu = dt^2 - a^2(t) \delta_{ij} dx^i dx^j = a^2(\eta) (d\eta^2 - \delta_{ij} dx^i dx^j) ,$$

with scale factor a , physical time t , conformal time η . Unless stated otherwise, f' denotes the derivative of f with respect to conformal time, while $\dot{f}$ denotes the derivative with respect to physical time.

Our convention for Fourier transformation in d dimensions is

$$X(x) = \int \frac{d^d k}{(2\pi)^{d/2}} \tilde{X}(k) e^{ikx}$$

$$\tilde{X}(k) = \int \frac{d^d x}{(2\pi)^{d/2}} X(x) e^{-ikx}$$

Throughout this thesis, we work in natural units, $c = \hbar = 1$, and define the reduced Planck mass as $M_{\text{P}} = 1/\sqrt{8\pi G} \simeq 2.435 \times 10^{18} \text{ GeV}$.

Contents

Introduction	1
I From inflation to axion inflation	7
1 Cosmic inflation	9
1.1 Prelude: Initial motivations for inflation	9
1.1.1 A brief reminder on FLRW cosmology	9
1.1.2 The problems of Big Bang cosmology	12
1.1.3 Accelerated expansion in the early Universe	15
1.2 Act I: Slow-roll inflation	16
1.2.1 Inflation driven by a scalar field	16
1.2.2 Models of slow-roll inflation	18
1.2.3 From inflation to radiation domination	23
1.3 Intermission: Quantum fields during inflation	24
1.3.1 Canonical quantization in an expanding spacetime	24
1.3.2 Super-horizon evolution and classicalization	26
1.3.3 The spectrum of a light scalar field during inflation	27
1.4 Act II: Primordial perturbations from slow-roll inflation	28
1.4.1 Preliminary considerations	29
1.4.2 Primordial scalar perturbations	30
1.4.3 Primordial tensor perturbations	32
1.5 Act III: The observable fingerprint of inflation	33
1.5.1 Super-horizon conservation: wavenumber as a clock variable	33
1.5.2 Slow-roll inflation and the anisotropies in the CMB	34
1.5.3 Signatures beyond the CMB	37
2 En route to axion inflation	41
2.1 A detour to axions	41
2.1.1 The non-trivial vacuum structure of Yang–Mills theory	43
2.1.2 Much ado about θ	45
2.1.3 The strong CP problem	48
2.1.4 Introducing the QCD axion	48
2.2 The phenomenology of axions	50
2.2.1 A closer look at the QCD axion	50
2.2.2 Recipe for a generic axion	51
2.3 The origin of axions	53
2.3.1 The Peccei–Quinn mechanism	53
2.3.2 Axions from string theory	55

3	Axion inflation	57
3.1	What is axion inflation?	57
3.1.1	Slow-roll with axions	58
3.1.2	Axion–gauge-field interactions during inflation	59
3.2	The model of Abelian axion inflation	60
3.2.1	The equations of motion	61
3.2.2	Gauge-field production and the tachyonic instability	62
3.2.3	Gauge-field backreaction	65
3.3	The dynamics of Abelian axion inflation	67
3.3.1	The regime of weak backreaction	68
3.3.2	The regime of strong backreaction	69
3.4	Implications of gauge-field production	71
3.4.1	Sourced primordial perturbations	71
3.4.2	Schwinger pair production	72
3.4.3	Other effects	73
II	The dynamics of axion inflation	75
4	The gradient-expansion formalism	77
4.1	The GEF formulation of axion inflation	78
4.1.1	Reformulating the equations of motion of axion inflation	78
4.1.2	Boundary terms	79
4.1.3	Approximate closing condition	81
4.2	Solving the GEF equations	82
4.2.1	The GEF algorithm	82
4.2.2	Slow-roll initial conditions: a standard benchmark for the GEF	83
4.3	Inhomogeneous backreaction and the GEF	86
4.4	Outlook: the role of the GEF within axion inflation	89
5	The stability of the Anber–Sorbo solution	91
5.1	What is the Anber–Sorbo solution?	92
5.2	Stability analysis with the GEF	93
5.2.1	The toy-model GEF	95
5.2.2	The linearized gradient-expansion formalism	96
5.3	Application to the Anber–Sorbo solution	97
5.3.1	Lyapunov exponents	98
5.3.2	The decay of the Anber–Sorbo solution	100
5.3.3	Including variations in the Hubble rate	103
5.4	More insights from the LGEF	105
5.5	Summary and conclusion	108
6	Revisiting the Schwinger effect during axion inflation	111
6.1	An introduction to the Schwinger effect	113
6.1.1	The Schwinger effect in Minkowski space	113
6.1.2	The Schwinger effect in de Sitter space	115
6.2	An effective treatment of the Schwinger effect in axion inflation	116
6.2.1	Axion inflation with charge carriers	117
6.2.2	The induced current	118

6.2.3	Damping in a conductive medium	122
6.2.4	Summary of the effective description	126
6.3	Implementation into the GEF	127
6.3.1	General formulation	128
6.3.2	Capturing scale-dependent effects	130
6.3.3	Truncation and error estimation	132
6.4	Case study: Axion inflation coupled to hypercharge	134
6.4.1	Modelling the production of charged particles	135
6.4.2	Implications of Schwinger damping	136
6.4.3	Fermion-tempered backreaction	141
6.5	Summary and conclusion	143
III	Primordial perturbations in axion inflation	147
7	Cosmological perturbation theory	149
7.1	A closer look at perturbation theory	149
7.1.1	Metric perturbations	150
7.1.2	Matter perturbations	152
7.1.3	The linearized Einstein equations	153
8	Gauge-field-induced scalar perturbations	157
8.1	The formalism	158
8.1.1	Scalar perturbations in axion inflation	159
8.1.2	Vacuum and induced curvature perturbations	164
8.1.3	The scalar power spectrum and bispectrum	166
8.2	First application: Axion inflation without backreaction	169
8.2.1	Taking inventory	170
8.2.2	Results	174
8.2.3	Discussion	178
8.3	Second application: Axion inflation with backreaction	180
8.3.1	Computing the scalar power spectrum using the GEF	180
8.3.2	Neglecting gauge-field backreaction	182
8.3.3	Including gauge-field backreaction	185
8.4	Summary and conclusion	189
9	Gravitational waves from axion inflation	191
9.1	Gauge-field-induced gravitational waves	192
9.1.1	Tensor perturbations in axion inflation	193
9.1.2	Calculation of the tensor power spectrum	194
9.1.3	From inflationary gravitational waves until today	197
9.2	Model specifications and constraints	199
9.2.1	Modelling of inflationary dynamics and CMB constraints	200
9.2.2	Gravitational-wave detection characteristics and constraints	202
9.3	Gravitational waves from pure axion inflation	203
9.3.1	A benchmark spectrum for pure axion inflation	203
9.3.2	Identifying relevant parameter-space regions	206
9.3.3	Results for pure axion inflation	207
9.3.4	Summary on constraints and detectability regions	211

9.4	Gravitational waves from fermionic axion inflation	213
9.4.1	A benchmark spectrum for fermionic axion inflation	213
9.4.2	Results for gauge-field-induced gravitational waves	214
9.4.3	Estimating fermion-induced gravitational waves	217
9.5	Summary and conclusion	219
Conclusion		223
A Whittaker functions		231
A.1	Properties of Whittaker functions	231
A.2	Whittaker functions in axion inflation	232
B Numerical determination of mode functions and power spectra		235
B.1	Solving a homogeneous mode equation	235
B.2	Retarded Green function	237
B.3	Induced scalar and tensor power spectra	239
C General expressions for electric and magnetic conductivities		241
D Equation of motion for ζ and scalar sources		243
D.1	Deriving the equation of motion for ζ	243
D.2	Expressions for the source functions K and L	246
Bibliography		249

Introduction

Over the last decades, increasingly precise measurements of the cosmic microwave background (CMB) have provided cosmologists with detailed insights into the physical state of our observable Universe in the earliest stages of cosmic expansion [12]. These measurements reveal that our primordial Universe had a remarkably simple and highly symmetric structure. For one, the uniformity of the CMB temperature across the observable sky indicates that the early Universe was nearly perfectly homogeneous and isotropic on large scales. Furthermore, small anisotropies in the temperature and polarization of the CMB provide evidence for the existence of small-scale inhomogeneities in the distribution of primordial matter [13], which appear to have followed a nearly scale-invariant and Gaussian distribution [14, 15].

This high degree of symmetry in the initial state of our Universe is consistent with our modern understanding of cosmology. In particular, assuming the symmetric initial distribution of primordial perturbations, the standard cosmological model, Lambda cold dark matter (Λ CDM), accurately describes their subsequent evolution into the large-scale structure observed today [16]. However, what Λ CDM does not provide is a mechanism for generating these initial conditions.

Instead, the prevailing idea in modern cosmology is that these conditions have their origin in a hypothesized phase of cosmic inflation [17–21] — a brief period of accelerated expansion preceding the standard cosmic history described by Λ CDM. During this inflationary phase, quantum fluctuations are rapidly stretched to cosmological size, where they are converted into classical curvature perturbations that then seed the observed CMB anisotropies [22–26].

Theoretical models that realize this idea are primarily built on the paradigm of slow-roll inflation [19, 20]. In its simplest form, slow-roll inflation is realized through a single, minimally coupled, scalar field. This inflaton field drives inflation by slowly rolling down its potential such that its dominant potential energy effectively behaves as a dynamically decaying cosmological constant. In the ensuing period of near-exponential expansion, the rapidly evolving gravitational background then generates curvature perturbations by exciting the vacuum fluctuations of both the spacetime metric and the inflaton field [22–26]. Similarly, this also amplifies tensor metric perturbations, which then generate a primordial spectrum of inflationary gravitational waves (GWs) [27–31].

Indeed, this simple dynamical setup is phenomenologically successful, as it predicts a spectrum of primordial perturbations that is nearly scale-invariant and Gaussian, consistent with modern CMB measurements [14, 32, 33]. Meanwhile, this agreement alone does not uniquely determine the underlying microphysical structure of inflationary models, and only serves to establish slow-roll inflation as an appealing dynamical mechanism whose observable predictions, i.e., an approximately scale-invariant spectrum of primordial curvature and tensor perturbations, serve as the baseline expectation for inflation.

At the same time, these predictions themselves can be sensitive to the assumptions regarding the microphysical structure of an inflation model. In particular, interactions between the inflaton and other matter fields, such as fermions, gauge fields, or additional scalars, can modify both the inflaton dynamics [34, 35] and affect the spectrum of primordial perturbations that are generated during inflation [36–40]. One important reason is that inflaton–matter interactions can be a channel for particle production, which can be understood as an immediate consequence of coupling a quantum field to the external time-dependent background of the evolving inflaton field [36, 37, 41–45].

If particle production occurs during inflation, it can leave observationally distinctive signatures from the standard predictions of slow-roll inflation. For instance, the direct coupling between the inflaton and a matter field can source inflaton perturbations, which then modify the spectrum of inflationary curvature perturbations. This can, for instance, lead to deviations from a Gaussian spectrum of curvature perturbations [46–48] that can be constrained by modern CMB data [15, 49]. Similarly, GWs can be sourced by particle production, which can enhance their amplitude and lead to spectral features that could be observed in the future through GW experiments [50–52].

Since signatures of particle production are a realistic detection prospect, it is therefore important to make accurate predictions for them. However, a consistent computation of these observable signatures can be challenging in certain models, as the direct interaction between the inflaton field and the matter field can also entail that particle production can effectively backreact on the dynamical evolution of the inflaton field itself. In systems where this backreaction is important, this can render the entire study of inflationary particle production highly complex, as the inflaton’s dynamics are not only linked to the expansion of spacetime itself, but also to the continued rate of particle creation. In this case, self-consistently accounting for this backreaction is one of the central challenges, because the overall efficiency of particle production during inflation can be modified by these backreaction effects, which can ultimately manifest itself in the amplitude and spectral features of the primordial perturbations that are generated by these fields.

The self-consistent description of this interplay between the dynamics of the inflaton field, particle production, and the effect it can have on observables, is an overarching theme of this thesis. The particular context in which this topic emerges in our work is inflation models where the inflaton field is a pseudoscalar that directly interacts with a species of Abelian gauge fields through its Chern–Simons operator

$$\mathcal{L} \supset \phi F_{\mu\nu} \tilde{F}^{\mu\nu}. \quad (1)$$

This interaction is widely studied in the context of inflation, because it is an efficient channel for producing primordial gauge fields [53, 54], which leads to inflationary observables that are highly sensitive to the dynamics of the inflaton field [55–58]. This is because gauge-field production in this model depends exponentially on the velocity of the inflaton field, which is consistently non-vanishing if one assumes inflation to be driven by slow-roll dynamics.

Models of inflation which incorporate an interaction as in Eq. (1) fall within the broader class of *axion inflation* models [54, 55, 59–64], which are motivated by considering a realistic embedding of an inflationary model within high-energy physics (cf. Ref [65] for a review). In models of axion inflation, one assumes that the inflaton field is identified with an axion, which are a class of pseudoscalar fields that are broadly studied in particle physics, cosmology, and astrophysics, for instance as dark matter candidates [66–68], or to resolve the so-called strong CP problem of quantum chromodynamics [69–72] (cf. Refs. [73–77] for reviews). Within the context of inflation, axions are considered appealing inflaton candidates, as they are naturally equipped with a shift symmetry which ensures the radiative stability of their potential. This property is highly desirable in an inflaton field, as conventional slow-roll inflation requires a nearly flat inflaton potential to ensure a lasting stage of inflation and to maintain consistency with CMB measurements. Within this context of axion inflation, the classically shift-symmetric interaction of the inflaton to gauge-fields, $\phi F \tilde{F}$, is one of the main interactions that is allowed from an effective-field-theory (EFT) perspective, and is, for instance, proposed as a compelling reheating channel in axion inflation [78–81]. In particular, within the context of this thesis, we will simply refer to the specific model of inflation that we are considering as “axion inflation”, although the precise high-energy physics embedding of this model proves to be less important for our discussion.

Instead, the work we pursue here aligns with the large body of literature which studies the phenomenological effects that are invariably linked to the efficient creation of gauge fields via the

interaction in Eq. (1). First among these observable implications are of course the effect of gauge-field production on the spectrum of primordial perturbations. Gauge-field-induced curvature and tensor perturbations can be greatly enhanced at small physical scales, corresponding to physical processes that take place near the end of inflation [82, 83]. In particular, because gauge-field production in axion inflation depends exponentially on the inflaton velocity, the resulting perturbation spectra can become strongly peaked at small scales. This can amplify the signal of inflationary GWs far beyond the usual amplitude of vacuum GWs from inflation, such that this signal may even be detectable by next-generation interferometer experiments [51, 82–88]. Similarly, the gauge-field-induced small-scale enhancements of the curvature power spectrum can be important, as they can favor the indirect production of primordial black holes (PBHs) [89, 90], which are often considered as a possible dark matter candidate [91–93].

Beyond this effect on perturbations, a feature which makes the inflaton–gauge-field interaction in Eq. (1) especially distinctive, is that it leads to the production of helical gauge fields [53, 54]. This is a consequence of coupling the inflaton to the parity-odd Chern–Simons operator: the non-zero inflaton velocity during inflation leads to a spontaneous breakdown of parity. The helical nature of these gauge fields also results in helical GW production, which is a unique signature that may be used to distinguish a possible signal of inflationary GWs from axion inflation from other primordial sources in future GW experiments [51, 58]. Other works study the production of gauge fields in axion inflation as a mechanism for generating helical primordial magnetic fields [53, 54, 94]. Such magnetic fields are one of the main possibilities that one can evoke to explain gamma-ray observations of blazars [95–97], which suggest the existence of large-scale magnetic fields permeating the voids between galaxies (see Refs. [98–100] for reviews). Finally, the decay of helical gauge fields after inflation has been proposed as a possible channel for introducing CP violation into the SM plasma that is needed for a variety of baryogenesis scenarios [101–105].

Irrespective of the precise physical effect, the study of gauge-field production in axion inflation requires the treatment of a complex dynamical system in which the inflaton field and gauge fields consistently exchange energy through their interaction. As long as the production of gauge fields remains small, these fields can be treated as effectively decoupled, and many observables, such as GW spectra, can be computed analytically assuming conventional slow-roll dynamics of the inflaton field [55–58]. However, many of the above-mentioned effects are most pronounced when gauge-field production is most effective, which coincides with strong gauge-field backreaction on the dynamics of the inflaton. In particular, gauge fields act as a source of friction for the inflaton [55], which can reduce its velocity. This itself reduces the efficiency of gauge-field production, which then again allows the inflaton to accelerate, resulting in a resurgence of gauge-field production. Collectively, this leads to a complex dynamical cycle where the two fields can rapidly exchange energy, and which ultimately needs to be studied numerically (see, e.g., Refs [106–109]).

In fact, the background dynamics of axion inflation become even more complex if one considers that gauge fields may also couple to other charged fields like fermions. In this case, the amplification of the gauge-field by the inflaton background can itself be a channel for particle production via the renowned Schwinger effect [110–112]; a long-standing prediction in quantum electrodynamics that a strong external electromagnetic field can lead to the creation of charged particle–anti-particle pairs from the vacuum (cf. Refs. [113–115] for reviews). This mechanism can also be efficient during inflation, provided sufficiently strong electromagnetic fields can be generated and maintained [116–119], as is precisely the case in axion inflation [120, 121]. This additional mode for particle production then only adds to the already rich phenomenology of axion inflation, as charged particles can also imprint themselves on the spectrum of inflationary perturbations [122, 123], they can be a means for reheating after inflation [124], or contribute to the density of dark matter in the present Universe [125, 126].

However, just as the inflaton and gauge fields can exchange energy via their interaction, the same is true for charged particles and gauge fields, which entails that Schwinger pair creation during axion inflation is not a stationary process: once charged particles are produced abundantly, they in turn affect the gauge fields themselves. This interaction can dampen the further generation of gauge fields, which itself limits the capacity for continued pair creation to take place [127–129]. The dynamics of this pair creation process need to be accounted for to accurately model the production of charged particles during axion inflation. At the same time, the effect of charge carriers is to hinder the production of gauge fields, this also reduces their potential for backreacting on the inflaton’s evolution [128]. Accounting for such effects is also important for established phenomenological predictions of axion inflation, as a reduced amplitude in gauge fields can, for instance, impact predictions for GW production [121] or magnetogenesis [127, 130].

This context then sets the stage for the work we present in this thesis, wherein we revisit several topics in the literature on axion inflation which invariably require a self-consistent treatment of the dynamical interplay between the inflaton field, the gauge field, and possible charged particles.

In a first part, our focus is entirely on dynamical effects in axion inflation. Our first analysis is regarding the *Anber–Sorbo* solution [55]: a proposed steady-state solution in the regime of strong gauge-field backreaction, where the inflaton’s velocity and gauge-field friction maintain a perfectly stable balance. Many early works on axion inflation assumed the stability of this solution, and relied on it to study strong-backreaction dynamics analytically. In our work, we substantiate claims that the Anber–Sorbo solution is unstable [131], while identifying previously unexplored dynamical regimes in which its stability may nevertheless be realized.

While the Anber–Sorbo solution exists in systems which assume the Schwinger effect is ineffective, our second analysis is instead dedicated to systems where it is relevant. Therein, we revisit a prevalent effective description of Schwinger pair production in axion inflation [104, 130, 132–134], which models the backreaction of charged particles as an effective damping medium. In our work, we resolve an existing ambiguity in this effective description regarding the precise modelling of this conducting medium. Moreover, we propose that gauge-field damping via charged particles should be modelled as a scale-dependent effect, in order to reflect that Schwinger pair creation proceeds on finite distance scales. Collectively, these results lead to a revised effective description of Schwinger damping that can be used in future applications.

Following our discussion of background dynamics, the second part of our analyses instead aims at studying gauge fields as a source primordial perturbations. Here, we reassess existing calculations of gauge-field-induced curvature perturbations to account for the fact that gauge fields may not only source inflaton perturbations but also metric perturbations. Our results serve to clarify that gauge-field-induced metric perturbations, however, appear to be a subdominant effect. Moreover, we perform a comprehensive parameter space study of GW production to reassess their detectability, both for models in which the Schwinger effect is ineffective, and for those where it is efficient. One particularly important result of this work is that we establish that, even when the Schwinger effect is effective at damping gauge-field production, GWs in axion inflation can still be a realistic observational target for next-generation ground-based or space-born GW interferometers.

The connecting thread that is interwoven with these four analyses, is the numerical method that we employ to treat the dynamics of axion inflation: the *gradient-expansion formalism* (GEF) [128, 135], which is a semi-analytical method that we rely on throughout this thesis to describe gauge-field production and backreaction during inflation. Starting from the original GEF proposed in Ref. [128], we continuously adapt, modify, and expand upon this method to suit our particular purposes. Taken together, these changes amount to the development of the GEF into a more comprehensive numerical framework which can be used to self-consistently determine the inflationary background evolution in axion inflation, and use this information to determine the effects on the

spectra of primordial perturbations. The essence of this numerical framework is the Python code, the *Gradient Expansion Formalism Factory*, or GEF [7], which has been developed alongside this thesis to obtain many of the results we present here, and which we have made publicly available for future studies of the rich phenomenology of axion inflation.

The discussion in this thesis is then separated into three parts, each consisting of three chapters. Part I provides the necessary theoretical background for our discussion of axion inflation. We start in chapter 1, wherein we present the basics of cosmic inflation. Chapter 2 gives a brief introduction to axions, in particular in the context of particle physics. Then, chapter 3 is dedicated to the synthesis of these two topics in the form of axion inflation. In particular, this last chapter is where we discuss the theoretical description and the phenomenology associated with the production of gauge fields in axion inflation. In part II our attention is entirely focussed to the discussion of the complex background dynamics in axion inflation models. We begin in chapter 4 with a broad introduction to the GEF, and position it within the context of other numerical methods for studying axion inflation. In chapter 5, we present the GEF's first concrete application, in that we use it to study the stability of the Anber–Sorbo solution, whereas chapter 6 is dedicated to the discussion of the Schwinger effect in axion inflation. Finally, part III is where we turn our attention towards primordial perturbations. We begin in chapter 7 by reviewing the formalism of cosmological perturbation theory with an emphasis on sourced metric and matter perturbations. In chapter 8, we discuss the relevance of gauge-field-induced scalar metric perturbations in axion inflation. Finally, chapter 9 is dedicated to the discussion of GWs in axion inflation, which is where we highlight the results of our comprehensive study regarding their detectability.

Part I

From inflation to axion inflation

Chapter 1

Cosmic inflation

In this chapter, we review the theoretical framework of cosmic inflation that forms the foundation for our analysis in this thesis. Our aim is to establish the necessary physics context and quantum-field theory (QFT) tools that we need for our later discussion of axion inflation.

We begin in Sec. 1.1 by reviewing the basic description of an expanding spacetime, and then follow the historical motivation that led to the idea of inflation as a period of early accelerated expansion. In Sec. 1.2, we introduce the slow-roll paradigm and discuss how accelerated expansion can be realized through the dynamics of a scalar inflaton field. Having established this basic dynamical model of inflation, we then discuss how perturbations are generated during inflation. In Sec. 1.3, we examine the behavior of quantum fields in an inflationary spacetime by means of a toy example. We then build on this discussion in Sec. 1.4, where we review how primordial curvature and tensor perturbations are generated during slow-roll inflation. Finally, in Sec. 1.5, we elaborate on how these perturbations may manifest themselves as anisotropies in the cosmic microwave background (CMB), what the current observational constraints on slow-roll inflation are, and what observable signatures primordial perturbations may induce apart from the CMB.

1.1 Prelude: Initial motivations for inflation

From a modern perspective, inflation is often studied because it provides a natural mechanism for generating the primordial perturbations that we measure in the CMB. However, inflation also resolves conceptual fine-tuning issues that arise when extrapolating the standard cosmic history described by Λ CDM to early times, before the emission of the CMB. Indeed, the original proposal of inflation by GUTH [18] was made precisely to address these problems of the hot Big Bang model.

For the purposes of establishing the theoretical formulation of inflation, it proves to be instructive to first discuss how inflation resolves these conceptual fine-tuning problems as this discussion naturally permits us to establish fundamental concepts when discussing an inflationary spacetime. To do so, we first give a short reminder regarding the basic theoretical description of an expanding spacetime in general relativity (GR) in terms of the homogeneous and isotropic *Friedmann–Lemaître–Robertson–Walker* (FLRW) metric. Following this, we then discuss the two main problems of the hot Big Bang model, namely the *horizon and flatness problems*, and argue why a period of early accelerated expansion like inflation resolves them.

The arguments we present here are thoroughly presented in many textbooks on cosmology, with our discussion representing a synthesis of Refs. [136–138].

1.1.1 A brief reminder on FLRW cosmology

On sufficiently large scales, observations of the large-scale structure of our Universe or the CMB suggest that our Universe can be accurately described as a statistically homogeneous and isotropic

spacetime. In GR, the expansion of such a spacetime is described by the FLRW metric,

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu = dt^2 - a^2(t) \left(\frac{dr^2}{1 - Kr^2} + r^2 [d\theta^2 + \sin^2 \theta d\phi^2] \right). \quad (1.1)$$

Here, $g_{\mu\nu}$ denotes the spacetime metric coefficients with the coordinate t denoting cosmic time, while r , ϕ , and θ denote spatial radial and angular coordinates. The constant K parametrizes the curvature of the spacelike hypersurfaces, such that $K > 0$ corresponds to positive curvature, $K < 0$ to negative curvature and $K = 0$ entails spatial flatness.¹ Modern measurements of the CMB anisotropies indicate that our Universe is consistent with a vanishing spatial curvature [16], and we therefore assume that we can set $K = 0$ for much of this thesis. But, for the time-being, we keep K general, as its value is important for the flatness problem. Finally, the spatial coordinates (r, θ, ϕ) in Eq. (1.1) are defined to be *comoving* with respect to the time-dependent scale factor a . In the case of vanishing spatial curvature, $K = 0$, this entails that, e.g., physical radial distances may be obtained by multiplication with the scale factor, $d_{\text{phys}} = a(t)r$.

The rate of change of the scale factor defines the Hubble rate, which we denote by

$$H \equiv \frac{\dot{a}}{a}. \quad (1.2)$$

Its evolution directly follows from Einstein field equations, which generally describes the dynamical evolution of the spacetime metric $g_{\mu\nu}$ in GR,

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = \frac{1}{M_{\text{P}}^2} T_{\mu\nu}. \quad (1.3)$$

Here, the Ricci tensor, $R_{\mu\nu}$, and the Ricci scalar, $R = g^{\mu\nu} R_{\mu\nu}$, describe the intrinsic curvature of spacetime, the energy-momentum tensor $T_{\mu\nu}$ describes matter, and $M_{\text{P}} = (8\pi G)^{-1/2}$ is the reduced Planck mass. Consistency with the assumption of a homogeneous and isotropic spacetime demands that the energy-momentum tensor is in the form of an ideal fluid,

$$T^\mu{}_\nu = (\rho + p)u^\mu u_\nu - p\delta^\mu{}_\nu, \quad (1.4)$$

with ρ being a homogeneous energy density of matter, p is its pressure, and u^μ is the four-velocity of a comoving observer. Einstein equations then reduce to the Friedmann equations

$$H^2 = \frac{\rho}{3M_{\text{P}}^2} - \frac{K}{a^2}, \quad (1.5a)$$

$$\frac{\ddot{a}}{a} = -\frac{1}{6M_{\text{P}}^2}(\rho + 3p). \quad (1.5b)$$

In particular, the first Friedmann equation (1.5a) is sufficient to describe the evolution of the scale factor $a(t)$. However, the second Friedmann equation (1.5b) is useful when discussing inflation, because it establishes a criterion for an accelerating expansion, $\rho + 3p < 0$, and a decelerating expansion, $\rho + 3p > 0$.

¹Often, one defines the FLRW metric in Eq. (1.1) such that the constant K takes values 1, -1 or 0. In this convention, the coordinate r would be dimensionless, and the scale factor a would have units of length. Our convention differs from this, as we define $K = 1/R_K^2$, $-1/R_K^2$ or 0 with R_K an arbitrary length scale describing curvature. This means that, in our notation, r has units of length, and a is dimensionless. Starting from Sec. 1.2 onwards, we consistently work with $K = 0$, and the value of R_K will be irrelevant for us.

The evolution of ρ is determined by covariant energy conservation, $\nabla_\mu T^\mu_\nu = 0$, which gives the continuity equation

$$\dot{\rho} + 3H(\rho + p) = 0, \quad (1.6)$$

For a constant equation-of-state parameter $w = p/\rho$, this equation may be directly integrated to yield the well-known scaling

$$\rho \propto a^{-3(1+w)}. \quad (1.7)$$

Typical equations of state in cosmology are $w = 1/3$ for relativistic matter (radiation), $w = 0$ for collisionless non-relativistic matter (dust), and $w = -1$, which corresponds to a cosmological constant. In particular, from Eq. (1.5b), we can note that accelerated cosmic expansion is driven by a fluid with $w < -1/3$, while decelerating expansion corresponds to $w > -1/3$. Therefore, out of these three standard fluids, only a cosmological constant drives a period of accelerated expansion.

In a spacetime with vanishing spatial curvature, $K = 0$, one may integrate the first Friedmann equation (1.5a) exactly to find

$$a(t) = \left(\frac{t}{t_0}\right)^{2/3(1+w)}, \quad H(t) = \frac{2}{3(1+w)t}, \quad (w > -1), \quad (1.8a)$$

$$a(t) = e^{H(t-t_0)}, \quad H(t) = \text{const.}, \quad (w = -1). \quad (1.8b)$$

Here, t_0 denotes a reference time at which $a(t_0) = 1$.² From Eq. (1.8) we can also note that a flat FLRW spacetime with only a cosmological constant is exponentially expanding with a constant Hubble rate, H . Such a spacetime is referred to as *de Sitter space*, and will be of particular importance to the discussion of inflation.

The current standard model of cosmology, Λ CDM, assumes that cosmic expansion is driven by the three energy-density constituents: radiation, ρ_r , matter, ρ_m , and a cosmological constant Λ with an energy density given by $\rho_\Lambda = \Lambda M_{\text{Pl}}^2$. Current CMB measurements from the Planck satellite are consistent with a spatially flat spacetime ($K = 0$), with our present-day Universe being composed of approximately 0.01% relativistic matter consisting of photons and neutrinos, 32% non-relativistic matter which further separates into 27% cold dark matter and 5% baryonic matter, while the remaining 68% are entirely attributed to dark energy [16]. Since the energy density of radiation scales as $\rho_r \propto a^{-4}$, that of dust scales as $\rho_r \propto a^{-3}$, and dark energy has a constant energy density, one can extrapolate backwards in time to conclude that the cosmic expansion history described by Λ CDM is one in which an initial era of radiation domination is followed by an era of matter domination, and where we are presently entering an era where the cosmological constant Λ is dominating the energy density of the Universe.

This expansion history can also be conveniently expressed in terms of the temperature of radiation in the Universe, T . From the viewpoint of thermodynamics, the energy density of radiation can be written as

$$\rho_r(T) = \frac{\pi^2}{30} g_*(T) T^4, \quad (1.9)$$

where $g_*(T)$ describes the effective number of relativistic degrees of freedom that are coupled to the thermal plasma described by ρ_r . From the equation of state of radiation, $p(T) = \rho(T)/3$, one may then infer that the entropy density of radiation behaves as

$$s(T) = \frac{2\pi^2}{45} g_{*,S}(T) T^3, \quad (1.10)$$

²A common convention is to define $a(t_0) = 1$ to correspond to the present day.

where $g_{*,S}(T)$ describes the effective number of relativistic degrees of freedom in entropy.³ Assuming that total comoving entropy is conserved, $S = a^3 s = \text{const.}$, Eq. (1.10) implies that the scale factor at two given moments of time, t_1 , and t_2 , is related to temperature via⁴

$$\frac{a_1}{a_2} = \frac{T_2}{T_1} \left(\frac{g_{*,S}(T_2)}{g_{*,S}(T_1)} \right)^{1/3}. \quad (1.11)$$

This approximately implies $T \propto a^{-1}$, such that we may equivalently describe cosmic evolution in terms of temperature instead of time.

1.1.2 The problems of Big Bang cosmology

Having reestablished these basics of FLRW cosmology, we can now turn towards the discussion of the flatness and horizon problem.

As we shall see in short, both of these problems arise in the context of the hot Big Bang model because it assumes that the early periods of cosmic expansion are described purely by decelerating expansion, i.e., first in the radiation-dominated era, and then in the matter dominated era. Fundamentally, the two problems both relate to the fact that the *comoving Hubble radius* $(aH)^{-1}$ is growing during a period of decelerating expansion, or, alternatively, that the *comoving Hubble scale* aH is shrinking. This can be deduced from Eq. (1.5b), which implies that

$$\frac{d}{dt}(aH)^{-1} = -\frac{1}{a} \left(1 + \frac{\dot{H}}{H^2} \right) = -\frac{1}{aH} \frac{\ddot{a}}{a}. \quad (1.12)$$

In both cases, the resolution will be to assume a period of accelerated expansion prior to radiation domination, i.e., inflation, in which the comoving Hubble radius shrinks and aH grows.

The horizon problem

The horizon problem ultimately amounts to a question of causality in FLRW spacetime. In a decelerating flat FLRW spacetime, such as the one described by Λ CDM, there is an initial singularity at time $t_i = 0$ where the scale factor $a(t_i) = 0$ vanishes, i.e., the usual Big Bang singularity. This means, information that is exchanged with the speed of light can only ever have traversed a finite distance until it has reached us today. The horizon problem notes that this can amount to a fine-tuning problem in explaining how our observable Universe may have attained a state of nearly perfect large-scale homogeneity. Extrapolating the cosmic history of Λ CDM backwards in time, our current observable Universe would consist of thousands of local patches of spacetime which would have been causally disconnected for much of our cosmic history (see, e.g., Ref. [137] for an explicit estimate of this statement). In particular, photons in the CMB would not have had sufficient time to equilibrate, and only extremely fine-tuned initial conditions would have allowed for the CMB photons to have a constant temperature of $T_0 = 2.7255K$ [139] across the entire sky, as we observe it today.

To better understand why this problem arises for decelerating expansion, and how a period of early accelerated expansion resolves it, we first need some basic aspects concerning the causal structure of FLRW spacetime. To this end, it is useful to define the conformal time coordinate

³The subtle difference between $g_*(T)$ and $g_{*,S}(T)$ relates to the fact that neutrinos have a slightly lower temperature than photons after electron–positron annihilation which occurs at a temperature $T \lesssim 0.5 \text{ MeV}$.

⁴We shall often denote the scale-factor at time X as $a_X = a(t_X)$, and will do the same for the temperature T and the Hubble rate H .

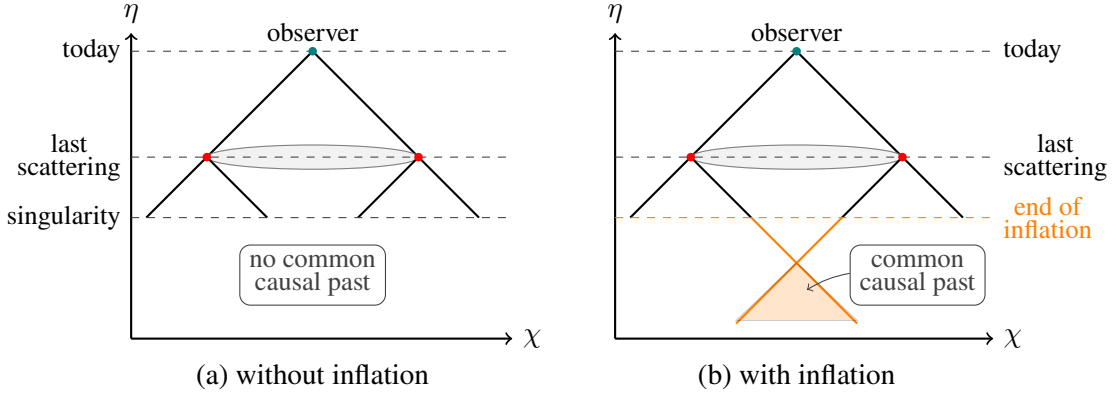


FIGURE 1.1: A schematic illustration how inflation resolves the horizon problem. Panel (a): without inflation, the surface of last scattering in our past light cone consists of patches which do not share a common causal past. Panel (b): with inflation, separated regions on the last-scattering surface have been in causal contact during inflation. *This figure is inspired from Ref. [138].*

$\eta = \int dt/a$, in terms of which the FLRW metric becomes conformal,

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu = a^2(\eta) (d\eta^2 - \{d\chi^2 + \chi^2 [d\theta^2 + \sin^2 \theta d\phi^2]\}) . \quad (1.13)$$

Here, we have set $K = 0$ for simplicity. Writing the metric in conformal time permits a simple study of radial light-like geodesics, as they are given by $d\chi/d\eta = \pm 1$. This way, one may deduce the existence of two complementary horizon scales, the *comoving particle horizon*, χ_p , and the *comoving event horizon*, χ_e . The two horizons are such that two events separated by $\chi > \chi_e(t)$ at time t will never reach causal contact, while events separated by $\chi > \chi_p(t)$ at time t have never been in causal contact at any earlier times.

Relevant for the horizon problem is the particle horizon, which is explicitly given by

$$\chi_p \equiv \eta - \eta_i = \int_{t_i}^t \frac{dt}{a(t)} = \int_{a_i}^{a(t)} \frac{da}{a} \frac{1}{aH(a)} , \quad (1.14)$$

with t_i and η_i denoting the initial time of cosmic expansion. In a decelerating flat spacetime, these can be defined to be the singularity at $t_i = 0$ corresponding to $\eta_i = 0$. From the second line, we can note that $\chi_p(t)$ for a decelerating spacetime is essentially dominated by the upper integration boundary in Eq. (1.14), i.e., $\chi_p(t) \simeq 1/(a(t)H(t))$,⁵ as the comoving Hubble radius, $(aH)^{-1}$, grows in this case. Therefore, the particle horizon in a decelerating spacetime remains finite at any time $t > t_i$, which then gives rise to the horizon problem.

The opposite is true in an accelerating spacetime. Here, $(aH)^{-1}$ shrinks, and we have that $\chi_p(t) \simeq (a_i H_i)^{-1}$ as $a_i H_i \ll a(t)H(t)$ for late enough times, t . In particular, for a sufficiently long inflationary period, we have that the particle horizon at time t can be essentially infinite: $\chi_p(t) = \eta - \eta_i \simeq (a_i H_i)^{-1} \rightarrow \infty$, for $a_i \rightarrow 0$. Alternatively, we can view this as the inflationary period extending the spacetime to earlier conformal times past $\eta = 0$: if inflation lasts sufficiently long, one has $\chi_p(t) = \eta - \eta_i \simeq (a_i H_i)^{-1} \rightarrow \infty$ such that $\eta_i \rightarrow -\infty$ is possible. In this way, apparently causally disconnected patches of spacetime can actually have shared a common causal past during inflation at negative conformal time. An illustration of this concept is given in Fig. 1.1.

⁵Because of this relationship, the comoving Hubble radius is also referred to as the *comoving Hubble horizon*.

The flatness problem

The flatness problem is similar to the horizon problem in that it again amounts to an issue of fine-tuned initial conditions. This time, the fine-tuning relates to the difficulty in maintaining a Universe which is nearly perfectly spatially flat, assuming that its expansion had been decelerating starting from the Big Bang singularity onward.

To explore the problem and its resolution through inflation, it is useful to define the density parameter for curvature

$$\Omega_K \equiv -\frac{K}{(aH)^2} = 1 - \sum_X \frac{\rho_X}{3H^2 M_{\text{P}}^2} \quad \text{for } X = m, r, \Lambda. \quad (1.15)$$

Looking back at Eq. (1.5a), we can note that Ω_K amounts to a measure of the deviation from spatial flatness throughout cosmic expansion. As previously discussed, CMB measurements can be used to infer that our current Universe is consistent with spatial flatness, with e.g., Planck data finding the present-day value of Ω_K to be, $\Omega_{K,0} = 0.0007 \pm 0.0037$ [140].

We can note that $\Omega_K \propto (aH)^{-2}$. During a period of decelerating expansion, $|\Omega_K|$ therefore grows as a function of time, such that its value in a previous era of cosmic history had to be tiny to be consistent with zero, as we infer it today. As an example, we can estimate Ω_K at the relevant time for Big Bang nucleosynthesis (BBN), which requires

$$|\Omega_K(t_{\text{BBN}})| = \left(\frac{a_0 H_0}{a_{\text{BBN}} H_{\text{BBN}}} \right)^2 |\Omega_{K,0}| \simeq \left(\frac{T_{\text{BBN}}}{T_0} \right)^2 \left(\frac{t_{\text{BBN}}}{t_0} \right)^2 |\Omega_{K,0}| \lesssim 10^{-18} \quad (1.16)$$

to be consistent with our current measurements. For this estimate, we used that $T_{\text{BBN}} \simeq 1$ MeV, $t_{\text{BBN}} \simeq 1$ s, and the current age of our Universe, $t_0 \simeq 13.8$ Gyr [16]. This means that a high degree of fine-tuning is necessary to ensure that our Universe appears to us as consistent with spatial flatness. Of course, the problem can be avoided if Ω_K vanishes exactly. However, from a theoretical perspective, there is no clear reason why this should be the case.

Inflation offers an appealing dynamical explanation. If we assume an inflationary period starting at t_I and ending at t_F one finds that $|\Omega_K(t_F)|$ at the end of inflation can be tiny even if $|\Omega_K(t_I)| \sim \mathcal{O}(1)$. The reason is that the comoving Hubble scale grows during inflation, which entails that $|\Omega_K(t_F)| \ll 1$ as long as $(a_I H_I)/(a_F H_F) \ll 1$.

We can then use Eqs. (1.15) and (1.16) to estimate by how much the scale factor needs to change to resolve the flatness problem:

$$10^{-18} \gtrsim |\Omega_{\text{BBN}}| \simeq \left(\frac{a_I H_I}{a_{\text{BBN}} H_{\text{BBN}}} \right)^2 = \left(\frac{a_I}{a_F} \right)^2 \left(\frac{T_F}{T_{\text{BBN}}} \right)^2, \quad (1.17)$$

with T_F denoting the temperature at the end of inflation, for which we assume an instantaneous transition to radiation domination. Moreover, we assumed a roughly constant Hubble rate during inflation, $H_I \simeq H_F$ (see further below). This can also be expressed as a bound on the duration of inflation in terms of the logarithmic change in the scale factor, the so-called *number of e-folds* of inflation, N . From Eq. (1.17), we obtain that the number of total e -folds of inflation needs to be larger than

$$N_{\text{tot}} \equiv \log \left(\frac{a_F}{a_I} \right) \gtrsim 20 + \log \left(\frac{T_F}{T_{\text{BBN}}} \right) \simeq 61 + \log \left(\frac{T_F}{10^{15} \text{ GeV}} \right), \quad (1.18)$$

to resolve the flatness problem. Here, in the second line, we used some typical temperature scale for the end of inflation that we will infer later on in Sec. 1.5. This number of 60 e -folds will be a

typical benchmark for the duration of inflation that we shall frequently refer to in what follows.

The monopole problem

Besides the horizon and the flatness problem, a period of inflation can also resolve other conceptual problems of Big Bang cosmology. One notable example is the *monopole problem* which is located in the context of QFT and particle physics. Monopoles are zero-dimensional cosmic defects. Such defects are topologically stable field configurations that may form as a consequence of the spontaneous breakdown of local or global symmetries [141–143]. Besides monopoles, other defects are the one-dimensional cosmic strings, two-dimensional domain walls, or textures which are extended unstable configurations in three-dimensional space. An interested reader may find Ref. [144] as a classic textbook on cosmic defects enlightening.

Monopoles can generically form in the spontaneous symmetry breaking of some grand unified gauge group to the Standard Model group $SU(3)_c \times SU(2)_L \times U(1)_Y$. A typical temperature scale at which this is thought to occur is $T_M \sim 10^{15}$ GeV, which corresponds to an epoch that lies deep inside the radiation-dominated era. The monopole problem amounts to the fact that these stable defects essentially behave as a type of collisionless matter, $\rho_M \propto a^{-3}$. Their energy density may therefore overclose the Universe, as it can come to dominate over that of radiation. A possible solution to avoid the monopole problem is to assume inflation happens after their formation, which dilutes their abundance before radiation domination can take place [144].

1.1.3 Accelerated expansion in the early Universe

We can surmise that a period of accelerated expansion before the radiation-dominated era is well-motivated as a means for resolving the above problems. Our focus shall later on shift to inflation as a mechanism for generating primordial perturbations, but the above discussion has been useful to establish some basic concepts about an inflationary spacetime. Let us finalize these with some concluding remarks before we move on to the discussion of inflation in the context of slow-roll inflation.

We can note that the problems of Big Bang cosmology are only resolved if inflation lasts sufficiently long. Indeed, we shall see in Sec. 1.5 that this is also needed for inflation to explain the anisotropies in the CMB. Looking back at Eq. (1.12), one may note that inflation corresponds to a period where the Hubble rate varies slowly with respect to the rate of expansion, $-\dot{H} < H^2$. A lasting period of expansion can then be conveniently expressed in terms of the two parameters that quantify the rate of change of H ,⁶

$$\varepsilon \equiv -\frac{\dot{H}}{H^2}, \quad \kappa \equiv \frac{\dot{\varepsilon}}{H\varepsilon}. \quad (1.19)$$

These are defined such that $\varepsilon < 1$ implies accelerated expansion, while $|\kappa| \ll 1$ can ensure that a period of inflation lasts for a prolonged time. Looking at Eq. (1.8), a first naive guess for an inflationary period is that it is driven by a cosmological constant for which $\varepsilon = \kappa = 0$. However, such a period of inflation would never end, and there would be no transition to the radiation-dominated epoch. A good alternative is a period of expansion driven by the potential energy of a scalar field, which we discuss in the next section. Regardless, assuming $H \simeq \text{const.}$ will prove to be a good approximation for an inflationary spacetime, corrections to which are of the order of ε .

⁶Similar parameters encoding higher-order time-derivatives of the Hubble rate may also be defined, but we shall not rely on them here.

In particular, since de Sitter space corresponds to $H = \text{const.}$, we often refer to a spacetime with $H \simeq \text{const.}$ as *quasi-de Sitter space*.

With these preliminary considerations, we can then continue by discussing how inflation can be realized through a scalar field following the slow-roll paradigm.

1.2 Act I: Slow-roll inflation

The idea that a scalar field can drive a period of inflation was first realized by GUTH [18]. His proposal was to drive inflation by having a scalar field be initially trapped in a false vacuum state such that its potential energy effectively acts as cosmological constant. In this model, inflation then ends via quantum tunneling to the true vacuum. This phase transition proceeds via the formation of true-vacuum bubbles, which collide, and releases energy in the form of radiation, thus leading to the onset of radiation domination. However, this model suffers a “graceful exit” problem [145, 146]: either the rate of bubble nucleation cannot compete with the expansion of spacetime, and inflation never ends, or bubble nucleation is efficient, but the radiation which is released through bubble collisions is too inhomogeneous, in conflict with observations.

This ultimately led to the development of slow-roll inflation by LINDE [19] as well as ALBRECHT AND STEINHARDT [20] which forgoes the need for bubble nucleation to end inflation. Instead, slow-roll inflation proposes a dynamical process that ensures a dominant potential energy and a subdominant kinetic energy of the scalar field. In slow-roll inflation, the scalar field slowly rolls towards its potential minimum by having the rapid expansion of spacetime act as an effective source of friction. At the same time, any deviations from homogeneity and isotropy in the scalar field are erased by the expansion diluting preexisting inhomogeneities.

In what follows, we will discuss the mathematical description of this model, explore various realizations of slow-roll inflation, and discuss how the onset of radiation domination after inflation can take place in such a model. In this, we primarily follow Refs. [137, 138], with input from additional references as indicated.

1.2.1 Inflation driven by a scalar field

Our starting point is the action for a real scalar field ϕ with a generic potential $V(\phi)$ in a curved spacetime,

$$S = S_{\text{EH}} + S_\phi = \int d^4x \sqrt{-g} \left[-\frac{M_{\text{P}}^2}{2} R + \frac{1}{2} g^{\mu\nu} \partial_\mu \phi \partial_\nu \phi - V(\phi) \right], \quad (1.20)$$

where S_{EH} is the Einstein–Hilbert action, and S_ϕ is the action for the field ϕ .

To obtain the classical equations of motion we vary S with respect to the metric $g_{\mu\nu}$ and the scalar field ϕ . Variation with respect to the metric yields the Einstein field equations in Eq. (1.3) with the energy–momentum tensor given by

$$T_{\mu\nu} = \frac{2}{\sqrt{-g}} \frac{\delta S_\phi}{\delta g^{\mu\nu}} = \partial_\mu \phi \partial_\nu \phi - g_{\mu\nu} \left(\frac{1}{2} g^{\alpha\beta} \partial_\alpha \phi \partial_\beta \phi - V(\phi) \right). \quad (1.21)$$

Variation with respect to the scalar field results in the Klein–Gordon equation in a curved spacetime,

$$\frac{1}{\sqrt{-g}} \partial_\mu [\sqrt{-g} g^{\mu\nu} \partial_\nu \phi] + \frac{dV(\phi)}{d\phi} = 0. \quad (1.22)$$

Obviously, the coupled equations given by Eqs. (1.3), (1.21), and (1.22) permit many possible solutions. However, we are interested in those describing a homogeneous, isotropic and flat spacetime, i.e., for an FLRW metric with $K = 0$:

$$g_{\mu\nu}dx^\mu dx^\nu = dt^2 - a^2(t)\delta_{ij}dx^i dx^j = a^2(\eta) [d\eta^2 - a^2(t)\delta_{ij}dx^i dx^j] . \quad (1.23)$$

The evolution of the scale factor is then determined by the Friedmann equations (1.5) for $K = 0$ with the energy density ρ and pressure p given by

$$\rho \equiv T^0_0 = \frac{1}{2}\dot{\phi}^2 + \frac{1}{2a^2}(\nabla\phi)^2 + V(\phi) , \quad (1.24a)$$

$$p \equiv -\frac{1}{3}\delta^j_i T^i_j = \frac{1}{2}\dot{\phi}^2 - \frac{1}{6a^2}(\nabla\phi)^2 - V(\phi) , \quad (1.24b)$$

with $\nabla = (\partial_x, \partial_y, \partial_z)$. The Klein–Gordon equation in an expanding background is

$$\ddot{\phi} + 3H\dot{\phi} - \frac{1}{a^2}\Delta\phi + V_{,\phi}(\phi) = 0 , \quad (1.25)$$

where we denote $V_{,\phi} = dV/d\phi$ and $\Delta \equiv \nabla^2$.

Obviously, our assumption of a homogeneous spacetime is only justified if ρ and p are homogeneous and isotropic. This motivates a perturbative Ansatz, $\phi(t, \mathbf{x}) = \varphi(t) + \delta\phi(t, \mathbf{x})$, where $\varphi(t)$ is homogeneous and isotropic, and $\delta\phi(t, \mathbf{x})$ are small perturbations away from it.⁷ In particular, we view $\varphi(t)$ as the (time-dependent) vacuum expectation value of a quantum field, $\varphi(t) = \langle\phi(t, \mathbf{x})\rangle$, and $\delta\phi(t, \mathbf{x}) = \phi(t, \mathbf{x}) - \langle\phi(t, \mathbf{x})\rangle$ as the quantum fluctuations around it. The notion of vacuum we are using here will be elaborated upon in Sec. 1.3.

For the time being, we focus entirely on the classical background contribution, φ , and disregard the perturbation $\delta\phi$. The dynamics of the scalar field and the expansion of spacetime are then described by

$$H^2 = \frac{1}{3M_{\text{P}}^2} \left(\frac{1}{2}\dot{\varphi}^2 + V(\varphi) \right) , \quad (1.26a)$$

$$\ddot{\varphi} + 3H\dot{\varphi} + V_{,\varphi}(\varphi) = 0 . \quad (1.26b)$$

In the first equation, we see that the energy density in the scalar field separates into a potential and a kinetic contribution. Meanwhile, the expansion of the background enters the Klein–Gordon equation as a friction term, $3H\dot{\varphi}$.

It is instructive to express the acceleration of the scale factor in terms of φ ,

$$\ddot{a} = \frac{a}{3M_{\text{P}}^2} (V(\varphi) - \dot{\varphi}^2) . \quad (1.27)$$

This way, it is evident that a homogeneous scalar field φ can drive a period of accelerated expansion as long as its potential energy is dominant, $V(\varphi) > \dot{\varphi}^2$. To further explore the potential for accelerated expansion, we can express the parameters ε and κ in Eq. (1.19) in terms of φ , finding

$$\varepsilon = 3 \frac{\dot{\varphi}^2}{\dot{\varphi}^2 + 2V(\varphi)} , \quad \kappa = 2 \frac{\ddot{\varphi}}{H\dot{\varphi}} + 2\varepsilon . \quad (1.28)$$

Once again, we see that accelerated expansion, $\varepsilon < 1$, happens for a small inflaton velocity, $\dot{\varphi}$, compared to the potential, $V(\varphi)$. Also, one may note that maintaining inflation for a sufficient

⁷We use this notation consistently throughout this thesis: $\phi(t, \mathbf{x})$ is the inflaton field, $\varphi(t)$ is the classical homogeneous field, and $\delta\phi(t, \mathbf{x})$ are deviations from $\varphi(t)$.

time, $|\kappa| \ll 1$ requires a small scalar-field acceleration, $\ddot{\phi}$. The assumption that $|\ddot{\phi}| \ll H|\dot{\phi}|$, while $\dot{\phi}^2 \ll V(\phi)$ are collectively referred to as *slow-roll approximation*. Using them in Eq. (1.26) the approximation is expressed as

$$H^2 \simeq \frac{V(\phi)}{3M_{\text{P}}^2}, \quad 3H\dot{\phi} \simeq -V_{,\phi}(\phi). \quad (1.29)$$

A solution satisfying these conditions is called a *slow-roll solution*, which is an attractor solution to Eq. (1.26) [147–149]. We may imagine such slow-roll dynamics as the scalar field ϕ rolling down its potential, with cosmic expansion acting as a source of friction that maintains a nearly constant inflaton velocity. Meanwhile, inflationary expansion is driven entirely by $V(\phi)$, which remains nearly constant, and thus realizes a stage of quasi-exponential expansion. However, unlike true exponential expansion through a cosmological constant, the scalar-field's kinetic energy can eventually become large enough to compete with the potential energy, thus stopping accelerated expansion.

The conditions for slow-roll inflation can be entirely formulated as conditions on the scalar potential $V(\phi)$ by defining the *slow-roll parameters*,

$$\varepsilon_V \equiv \frac{M_{\text{P}}^2}{2} \left(\frac{V_{,\phi}}{V} \right)^2, \quad \eta_V \equiv M_{\text{P}}^2 \frac{V_{,\phi\phi}}{V}. \quad (1.30)$$

These two parameters relate to those in Eq. (1.19) as $\varepsilon_V \simeq \varepsilon$ and $\eta_V \simeq 2\varepsilon - \kappa/2$. Hence, inflation happens while $\varepsilon_V < 1$, and slow-roll inflation in particular occurs for $\varepsilon_V, |\eta_V| \ll 1$. Phrasing this condition differently, we find that slow-roll inflation can occur if the inflaton field is on a potential height $V(\phi)$ with a sufficiently small potential gradient: $|V_{,\phi}(\phi)| \ll \sqrt{2V(\phi)}/M_{\text{P}}$.

Finally, let us consider the condition that ensures a lasting duration of inflation in the slow-roll approximation. Assuming a field starting/ending at $\phi_{i/e}$ at times $t_{i/e}$, the number of e -folds elapsed during inflation can be expressed as

$$N = \int_{t_i}^{t_e} dt H(t) = \int_{\phi_i}^{\phi_e} d\phi \frac{H}{\dot{\phi}} \simeq \int_{\phi_i}^{\phi_e} d\phi \frac{\text{sgn}(-V_{,\phi})}{M_{\text{P}}\sqrt{2\varepsilon_V}}. \quad (1.31)$$

In this expression, the sign of the potential needs to be such that $N > 0$. Equation (1.31) makes it convincing that $N \gtrsim 60$ can feasibly be achieved if $\varepsilon_V \ll 1$, while $|\eta_V| \ll 1$ ensures that the first condition remains true for a sufficiently long time.

1.2.2 Models of slow-roll inflation

Now that we have established the basic formulation of slow-roll inflation, we next want to consider models in which these dynamics can be successfully realized. We start our discussion by considering two basic classes of potentials on which a stage of slow-roll inflation can take place. These are known as the *large-field and small-field models* of inflation (see, e.g., Ref. [137]). As a representation of these models, we will describe both in terms of simple power-law potentials. However, it should be noted that the landscape of possible inflationary potentials is far more vast than this, so much so, that an entire encyclopedia of such potentials has been compiled [150]. Following a slightly more detailed discussion of these two basic models, we then give a concise overview about other models of slow-roll inflation, including some that go beyond a single inflaton field on a flat potential.

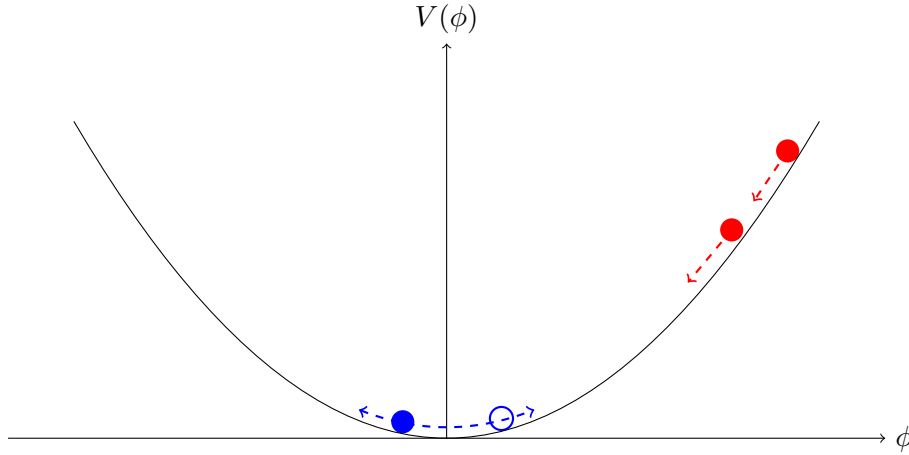


FIGURE 1.2: Illustration of large-field inflation. Inflation occurs for the field at large values (red). After inflation, the field oscillates around its minimum (blue). *This figure is reproduced from Ref. [137].*

Large-field models

Large-field models of inflation (historically: *chaotic inflation* [21]) are best exemplified by a simple monomial potential,

$$V(\phi) = g\phi^n, \quad n > 0, \quad (1.32)$$

where g is some coupling constant with the appropriate mass dimension to ensure that V has mass-dimension four. For these potential shapes, the slow-roll parameters in Eq. (1.30) are given by

$$\varepsilon_V = \frac{M_{\text{P}}^2 n^2}{2\phi^2}, \quad \eta_V \simeq \frac{n(n-1)M_{\text{P}}^2}{\phi^2}. \quad (1.33)$$

We can see that slow-roll inflation, $\varepsilon_V, |\eta_V| \ll 1$ occurs for super-Planckian field values, $\phi \gg nM_{\text{P}}/\sqrt{2}$, hence the name of these models. Inflation ends when $\phi_e \simeq nM_{\text{P}}/\sqrt{2}$, which is still at a Planckian distance from the origin of field space. The end of inflation coincides with $\dot{\phi}^2 \simeq V(\phi)$, and the remaining potential and kinetic energy then result in oscillations of the inflaton field around the potential minimum. An illustration of this evolution is given in Fig. 1.2.

For a power-law potential, we can directly compute the duration of inflation from Eq. (1.31) starting from some field value ϕ until $\phi_e \simeq nM_{\text{P}}/\sqrt{2}$,⁸

$$N(\phi) \simeq \frac{\phi^2}{2nM_{\text{P}}^2} - \frac{n}{4} \implies \frac{\phi}{M_{\text{P}}} \simeq \sqrt{2nN + \frac{n^2}{2}}. \quad (1.34)$$

Two standard benchmarks for this model are a quadratic and a quartic inflaton potential. Taking Eq. (1.34), one finds that 60 e -folds of inflation are realized for the inflaton field at

$$\phi \simeq \begin{cases} 15.55M_{\text{P}}, & \text{for } n = 2, \\ 22.09M_{\text{P}}, & \text{for } n = 4. \end{cases} \quad (1.35)$$

Seeing that inflation ends at $\phi_e \simeq M_{\text{P}}$, this explicitly shows the need for the inflaton to traverse a Planckian distance in field space in these models.

⁸For definiteness, we assume that $\phi > 0$, so the sign of $V_{,\phi}$ is positive.

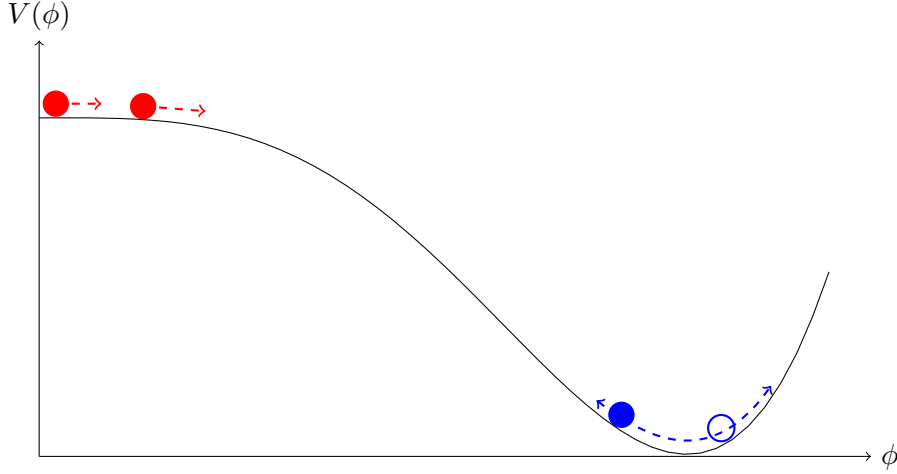


FIGURE 1.3: Illustration of small-field inflation. Inflation occurs for the field at small values (red). After inflation, the field oscillates around its minimum (blue). *This figure is reproduced from Ref. [137].*

Small-field models

As the name suggests, small-field models (historically: *new inflation* [19, 20]) do not require super-Planckian field excursions to realize slow-roll inflation. This is ensured by having inflation occur near a flat potential height. An illustration of this is given in Fig. 1.3.

To convey the idea of small-field inflation, it suffices to approximately write their potential near its maximum at $\phi = 0$ as

$$V(\phi) \simeq V_0 - g\phi^n. \quad (1.36)$$

For $g\phi^n \ll V_0$, the slow-roll parameters are given by

$$\varepsilon_V \simeq \frac{M_{\text{P}}^2}{2} \frac{g^2 n^2 \phi^{2n-2}}{V_0^2}, \quad \eta_V \simeq -\frac{g M_{\text{P}}^2 n(n-1) \phi^{n-2}}{V_0}, \quad (1.37)$$

implying the hierarchy $\varepsilon_V \ll |\eta_V|$ for $g\phi^n \ll V_0$, and inflation ends when $|\eta_V| \sim 1$.

Evidently, for $V_0 \ll g\phi^n$, the potential in Eq. (1.36) would be similar to large-field inflation. However, inflation occurs in a qualitatively different regime by demanding inflation end while $g\phi^n \ll V_0$. This restricts V_0 to

$$g\phi^n \ll V_0 \ll g M_{\text{P}}^n [n(n-1)]^{n/2}. \quad (1.38)$$

This condition, together with $|\eta_V| \ll 1$, implies that small-field inflation typically requires $n > 2$. Inflation can also occur in intermediate regimes, between small-field and large-field models. In this case, the condition in Eq. (1.38) can be relaxed. More generally, models where inflation occurs near a maximum in the scalar potential are referred to as *hilltop inflation* [151].

Note that, at larger field values, the potential $V(\phi)$ can significantly deviate from Eq. (1.36). After inflation, one usually assumes the inflaton reaches a minimum in $V(\phi)$, around which it then oscillates, just like in the case of large-field models (see Fig. 1.3).

The η -problem

One question that may arise when discussing inflaton potentials is how they originate in a high-energy theory which gives rise to the inflaton field. From the perspective of an effective field theory (EFT), the tree-level Lagrangian can in principle have corrections from non-renormalizable higher order terms suppressed by some UV scale Λ . Including these higher-order operators, the Lagrangian describing the inflaton can be written as (we implicitly assume a $\mathbb{Z}_2$ symmetry, $\phi \rightarrow -\phi$) [138]

$$\mathcal{L}_{\text{eff}}(\phi) = \frac{1}{2}(\partial\phi)^2 - V(\phi) - \sum_n c_n V(\phi) \left(\frac{\phi}{\Lambda}\right)^{2n} + \sum_n d_n \left(\frac{\partial\phi}{\Lambda^2}\right)^{2n}. \quad (1.39)$$

For $\phi < \Lambda$, this can be truncated at the lowest order in n , but even the dimension-six operator in this expression gives a non-negligible correction to the η_V slow-roll parameter,

$$\Delta\eta_V \simeq 2c_1 \left(\frac{M_{\text{P}}}{\Lambda}\right)^2, \quad (1.40)$$

which, for $c_1 = \mathcal{O}(1)$ and $\Lambda < M_{\text{P}}$ spoils the flatness of the potential needed for slow-roll dynamics, i.e., $|\eta_V| \ll 1$. A possibility to resolve this so-called η -problem are inflation models where the inflaton field enjoys an approximate shift symmetry, $\phi \rightarrow \phi + c$, which controls large UV corrections to the inflaton potential [59]. As mentioned in the introduction, this is the case for axion-like fields and serves as the original motivation for studying models of axion inflation (see chapter 3).

Alternative models

The large-field and small-field models we discussed above illustrate the basic idea behind a large class of conventional slow-roll models which simply assume a single inflaton field with slow-roll dynamics being regulated by the potential shape. However, there are many variants of slow-roll inflation that go beyond this idea. In what follows, we present a selection of such alternative models to give an impression of the variety of concepts. A detailed description of each scenario is beyond the scope of this thesis, and we refer to the respective references for more details.

Starobinsky inflation: STAROBINSKY [17] realized the first successful model of inflation by modifying gravity to allow quadratic terms in the Ricci Scalar, R , in the action for gravity,

$$S = -\frac{M_{\text{P}}^2}{2} \int d^4x \sqrt{-g} \left[R - \frac{R^2}{6M^2} \right]. \quad (1.41)$$

In such a modified model of Einstein gravity, gravity contains an additional scalar degree of freedom which can be identified as the inflaton field. This field can be made explicit via a transformation of the metric, and its potential is

$$V(\phi) = \frac{3}{4} M^2 M_{\text{P}}^2 \left(1 - e^{-\sqrt{\frac{2}{3}} \frac{\phi}{M_{\text{P}}}} \right)^2. \quad (1.42)$$

Ultimately, Starobinsky inflation therefore also amounts to a model of slow-roll inflation with a single inflaton field.

Hybrid inflation: This model realizes slow-roll inflation by introducing a second scalar field, χ (sometimes called the “waterfall field”) [152]. The main idea is to initially have χ stabilized at $\chi = 0$ and the inflaton field $\varphi = \langle \phi \rangle$ in a slow-roll potential, $U(\phi)$. However, the scalar potential features a saddle point. Once φ passes it, χ acquires a non-zero vacuum expectation value while the inflaton settles at $\varphi = 0$, thus ending inflation. For example, the scalar potential in hybrid inflation can be of the form

$$V(\phi, \chi) = \frac{1}{2}(g^2\phi^2 - \mu^2)\chi^2 + \frac{h}{4}\chi^4 + U(\phi) + V_0 \quad (1.43)$$

with $g, h > 0$ and μ a mass parameter. One advantage of this model is in the large variety of scalar potentials in which inflation can occur.

Multifield inflation: In a sense, hybrid inflation may be regarded as a particular scenario of multifield inflation. As the name suggests, these models feature multiple scalar fields that are active during inflation. There are two main varieties of multifield inflation: either slow-roll inflation proceeds along a trajectory through field space which is not aligned with a single scalar field, or the additional scalar fields are spectators to inflation (see, e.g., Ref. [153] for a review).

An example of the first type of multifield inflation is *assisted inflation* [154]. Therein multiple scalar fields φ_I can support inflation, even though the potential for each individual field is too steep. This is possible because Hubble friction is sensitive to the total energy density stored in all scalar fields. Typically, one can express these models as single-field variants, by identifying the direction in field space which determines the background expansion with the inflaton [155].

A common example for the second kind is the curvaton model [156–158]. Here, a heavy scalar (the inflaton) drives inflation, while the quantum fluctuations of a light spectator field (the curvaton) source the primordial perturbations needed to explain the CMB anisotropies (see Sec. 1.5).

K-flation: These models [159] differ substantially from the usual slow-roll paradigm in that they do not require the inflaton’s potential energy to drive inflation. Instead, one allows for a non-canonical kinetic term for the inflaton field, e.g.,

$$S_\phi = \int d^4x \sqrt{-g} P(\phi, (\partial\phi)^2), \quad (1.44)$$

where the function P admits an expansion of the kind

$$P(\phi, (\partial\phi)^2) = \frac{1}{2}K(\phi)(\partial\phi)^2 + \frac{1}{4}L(\phi)(\partial\phi)^4 + \dots \quad (1.45)$$

Such modified kinetic terms can, for instance, arise in the low-energy limit of string theories.

Warm inflation: All models we have discussed so far can be thought of as “cold inflation” models, in the sense that any particle interactions during inflation are regarded as being too inefficient to compete with Hubble expansion, such that radiative degrees do not thermalize. Warm inflation [160, 161] considers the opposite case, assuming that a thermal bath of temperature $T \gtrsim H$ can be sustained throughout inflation. The potential energy density of the inflaton still drives cosmic expansion, but the inflaton can now dissipate energy into its thermal environment. A simple understanding of these dynamics can be had by considering the coupled differential equations

$$\ddot{\varphi} + [3H + \Upsilon]\dot{\varphi} + V_{,\varphi} = 0, \quad (1.46a)$$

$$\dot{\rho}_r + 4H\rho_r = \Upsilon\dot{\varphi}^2. \quad (1.46b)$$

Here, $\Upsilon\dot{\phi}$ is a dissipative term and ρ_r is the energy density of radiation. Effectively, Υ gives an additional contribution to Hubble friction, which can support slow-roll inflation on potentials which would otherwise be too steep (c.f. Eq. (1.29)). In this way, warm inflation alleviates the restrictions of a shallow inflaton potential, and one forgoes the need to reheat the universe after inflation has ended [162]. Warm inflation also qualitatively differs in its experimentally testable predictions: unlike in cold inflation, the primordial perturbations in warm inflation are not sourced by quantum fluctuations but by classical thermal fluctuations instead [163]. For reviews on warm inflation, see Refs. [164–166].

1.2.3 From inflation to radiation domination

To close this introductory discussion of slow-roll inflation, we now consider how inflation ends and radiation domination can proceed. After (cold) inflation has ended, once $\dot{\phi}^2 \simeq V(\phi)$, the universe is cold and empty. Therefore, the inflaton field needs to transfer its energy to radiation such that the radiation-dominated era can take place. This process of creating a hot universe after inflation is called *reheating*. Here, we give only a brief overview of this topic in which we mainly follow Ref. [167]. For a more elaborate discussion of reheating there are dedicated reviews, such as Refs. [168–170].

To begin, we focus on the dynamical evolution of spacetime after inflation has ended, postponing the topic of energy transfer to later. For simplicity, we assume a single inflaton field with the potential shape around its minimum given by $V(\phi) \propto \phi^p$. Following the end of inflation, one can generically expect that the inflaton energy density is still dominant with respect to other matter components, the only difference is that accelerated expansion has stopped. This implies that the Hubble rate starts to decrease more rapidly, and Hubble friction for the inflaton field soon becomes small, $H \ll V_\phi/\dot{\phi}$, (c.f. Eq. (1.26)). Once this happens, the inflaton's dynamics are simply that of a harmonic oscillator, and the field starts to rapidly oscillate around its potential minimum. One can associate an average energy density with the inflaton oscillations, $\bar{\rho}_\phi$, with an effective equation of state parameter $\bar{w}_\phi = \bar{p}_\phi/\bar{\rho}_\phi$ where $\bar{p}_\phi$ is the averaged pressure over the oscillations. For $V(\phi) \propto \phi^p$ one can compute that this average equation of state parameter is given by $\bar{w}_\phi = (p-2)/(p+2)$ which follows from the virial theorem [171].

Assuming cosmic expansion during reheating is indeed driven by inflaton oscillations with a constant equation of state $\bar{w}_\phi$, one can estimate the duration of the reheating phase by considering that reheating ends when thermal radiation has reached $\rho_r(T_{\text{reh}}) = \bar{\rho}_\phi$ with T_{reh} the *reheating temperature*. Obviously, this requires the inflaton to lose energy into radiation, a discussion of which follows shortly. For now, we assume the energy transfer is only important in the final stages of reheating, such that expansion is driven by $\bar{\rho}_\phi \propto a^{-3(1+\bar{w}_\phi)}$. The duration of reheating in terms of e -folds is then easily found to be given by

$$N_{\text{reh}} \simeq \frac{1}{3(1+\bar{w}_\phi)} \log \left(\frac{90M_{\text{P}}^2 H_{\text{end}}^2}{\pi^2 g_*(T_{\text{reh}}) T_{\text{reh}}^4} \right), \quad (1.47)$$

where H_{end} is the Hubble rate at the end of inflation. A useful simplification that one often makes is *instantaneous reheating*, an immediate conversion of inflaton energy into radiation, i.e., $N_{\text{reh}} = 0$. In this case, the reheating temperature is related to H_{end} by

$$H_{\text{end}} = \frac{\pi T_{\text{reh}}^2}{3M_{\text{P}}} \sqrt{\frac{g_*(T_{\text{reh}})}{10}}, \quad (\text{instantaneous reheating}). \quad (1.48)$$

The two most common types of processes by which the inflaton field can transfer its energy into thermal radiation are referred to as *perturbative reheating* and *preheating*. As the name

suggests, perturbative reheating relates to the perturbative decay of inflaton particles into other particle species. Such reheating processes were proposed early on in the development of inflation, in particular in Refs. [172–174]. For scalars, three-scalar interactions are typically relevant, e.g., $g\phi\chi^2$ with ϕ the inflaton, and χ a second scalar field. For fermions, one can consider a Yukawa coupling, $\phi\bar{\psi}\psi$, where ψ denotes the fermion field.

Preheating concerns nonperturbative production processes wherein secondary particles are produced far from thermal equilibrium, and later decay to radiation. The first to study preheating processes were Refs. [41, 42], whose work was further developed by Refs. [43–45]. In these first models, the authors considered a quartic coupling, $g\chi^2\phi^2$, between the inflaton and a second light scalar field, χ . The production of scalar bosons then proceeds via a parametric resonance, as scalar-field fluctuations are effectively sourced by the oscillations of the classical background inflaton field, ϕ , leading to a large occupation number of bosons. Via the same mechanism of parametric resonance, inflaton couplings to fermions, e.g., Yukawa interactions, can also lead to efficient preheating, despite Pauli-blocking limiting the occupation numbers of fermions [175–179]. A different type of production mechanism takes place in models like hybrid inflation, where a second scalar descends from a local maximum as the inflaton field is oscillating. In this process, the mass of the second scalar field can become tachyonic, resulting in abundant particle production [180, 181]. A tachyonic instability also is the mechanism of gauge-field production in axion inflation, as we shall discuss in chapter 3.

Note that also more exotic scenarios of reheating exist, for example via Hawking evaporation of primordial black holes [182] (for more on primordial black holes, see Sec. 1.5).

1.3 Intermission: Quantum fields during inflation

The previous section has established the basic dynamical picture of slow-roll inflation up until the epoch of radiation domination. In the following sections, we want to consider how the vacuum fluctuations of quantum fields during inflation generate the primordial perturbations that are ultimately thought to leave observable imprints in our present-day Universe, most notably by sowing the seeds for the CMB anisotropies.

However, a full analysis of inflationary perturbations involves intrinsic subtleties related to perturbation theory in GR which are conflated with the discussion of QFT in an expanding background. We intend to separate these two topics by first discussing a simple, yet instructive, toy model example: a scalar field with a mass term during a phase of accelerated expansion. We will give a rather thorough presentation of this example, as it allows us to establish several important features of QFT in curved spacetime that are relevant when we describe the production of gauge fields during axion inflation in later parts of this thesis.

The discussion we present here is based on Ref. [183], with some additional insights taken from Ref. [138]. A classic textbook on QFT in curved spacetime is Ref. [184].

1.3.1 Canonical quantization in an expanding spacetime

To start, we take the action for a scalar field ϕ as in Eq. (1.20) with a potential $V(\phi) = m^2\phi^2/2$. Instead of first deriving the equations of motion, we immediately restrict the metric to that of flat FLRW space and obtain

$$S_\phi = \frac{1}{2} \int d\mathbf{x}^3 d\eta a^2 [\dot{\phi}^2 - (\nabla\phi)^2 - m^2 a^2 \phi^2] , \quad (1.49)$$

with η denoting conformal time. This closely resembles the action for a scalar field in flat spacetime, except for the appearance of the scale factor, a . To make the resemblance more pronounced, we introduce an auxiliary field, $\chi = a\phi$ whose action is

$$S_\chi = \frac{1}{2} \int d\mathbf{x}^3 d\eta \left[\chi'^2 - (\nabla\chi)^2 - \left(m^2 a^2 - \frac{a''}{a} \right) \chi^2 \right]. \quad (1.50)$$

Varying this action with respect to χ , we get the equation of motion (EOM)

$$\chi'' - \Delta\chi + \left(m^2 a^2 - \frac{a''}{a} \right) \chi = 0, \quad (1.51)$$

which resembles the well-known Klein–Gordon equation in flat spacetime except for the time dependent mass, $m_{\text{eff}}^2(\eta) = m^2 a^2 - a''/a$.

To quantize χ , one follows the usual steps; identifying its canonically conjugate momentum $\pi = \chi'$, and promoting both to operators by imposing the equal-time commutation relations,

$$[\chi(\eta, \mathbf{x}), \pi(\eta, \mathbf{y})] = i\delta^{(3)}(\mathbf{x} - \mathbf{y}), \quad (1.52a)$$

$$[\chi(\eta, \mathbf{x}), \chi(\eta, \mathbf{y})] = [\pi(\eta, \mathbf{x}), \pi(\eta, \mathbf{y})] = 0. \quad (1.52b)$$

Next, we would like to go to Fourier space, and introduce creation and annihilation operators just as we would for flat spacetime. However, we need to take into account the time-dependent mass term. Therefore, consider instead the following Ansatz,⁹

$$\chi(\eta, \mathbf{x}) = \int \frac{d^3\mathbf{k}}{(2\pi)^{3/2}} \left[v_k(\eta) \hat{a}_{\mathbf{k}} + v_k^*(\eta) \hat{a}_{-\mathbf{k}}^\dagger \right] e^{i\mathbf{k}\cdot\mathbf{x}}, \quad (1.53)$$

where “*” denotes complex conjugation and “†” the hermitian conjugate. The expression is written such that χ is real. The time-dependent coefficients, v_k are called *mode functions*. The EOM for χ implies that v_k must solve,

$$v_k'' + \left(k^2 + a^2 m^2 - \frac{a''}{a} \right) v_k = 0. \quad (1.54)$$

Turning to the operators $\hat{a}_{\mathbf{k}}^\dagger$ and $\hat{a}_{\mathbf{k}}$, we would like to interpret them as creation and annihilation operators, so they should obey the canonical commutation relations,

$$[\hat{a}_{\mathbf{k}}, \hat{a}_{\mathbf{k}'}^\dagger] = \delta^{(3)}(\mathbf{k} - \mathbf{k}'), \quad [\hat{a}_{\mathbf{k}}, \hat{a}_{\mathbf{k}'}] = [\hat{a}_{\mathbf{k}}^\dagger, \hat{a}_{\mathbf{k}'}^\dagger] = 0. \quad (1.55)$$

This would allow for the definition of a vacuum state $|0\rangle_a$ via $\hat{a}_{\mathbf{k}} |0\rangle_a = 0$. We thus want $\hat{a}_{\mathbf{k}}$ and $\hat{a}_{\mathbf{k}}^\dagger$ to satisfy Eq. (1.55), but, at the same time, χ and $\pi = \chi'$ need to obey the equal-time commutation relation in Eq. (1.52). The two conditions impose a constraint on the mode functions v_k and v_k^* :

$$W[v_k, v_k^*] \equiv v_k' v_k^* - (v_k^*)' v_k = i. \quad (1.56)$$

Equation (1.56) is equivalent to stating that v_k and v_k^* need to be linearly independent solutions to Eq. (1.54) (with the correct normalization). However, this condition is not sufficient to uniquely define v_k and v_k^* . Alternative mode functions satisfying Eq. (1.56) can be constructed by performing

⁹For notational simplicity, we henceforth denote only the creation and annihilation operators of fields with hats, while the operator nature of all other quantities is left implicit.

a Bogolyubov transformation:

$$u_k = \alpha_k v_k + \beta_k v_k^*, \quad u_k^* = \alpha_k^* v_k^* + \beta_k^* v_k \quad \text{with} \quad |\alpha_k|^2 - |\beta_k|^2 = 1. \quad (1.57)$$

It is easy to show that then $W[u_k, u_k^*] = i$. Expressing Eq. (1.53) in terms of the new mode functions u_k , we can see that the Bogolyubov transformation also leads us to define new operators,

$$\hat{b}_{\mathbf{k}} = \alpha_{\mathbf{k}}^* \hat{a}_{-\mathbf{k}} - \beta_{\mathbf{k}}^* \hat{a}_{\mathbf{k}}^\dagger, \quad \hat{b}_{\mathbf{k}}^\dagger = \alpha_{\mathbf{k}} \hat{a}_{\mathbf{k}}^\dagger - \beta_{\mathbf{k}} \hat{a}_{-\mathbf{k}}. \quad (1.58)$$

Because u_k satisfies $W[u_k, u_k^*] = i$, these new operators also obey canonical commutation relations as in Eq. (1.55), and we can define a vacuum state for them via $\hat{b}_{\mathbf{k}} |0\rangle_b = 0$. Since $\hat{b}_{\mathbf{k}}$ and $\hat{b}_{\mathbf{k}}^\dagger$ are linear combinations of $\hat{a}_{\mathbf{k}}$ and $\hat{a}_{\mathbf{k}}^\dagger$, this vacuum state is clearly distinct from $|0\rangle_a$, unless $\beta_{\mathbf{k}} = 0$. Thus, a transformation such as in Eq. (1.57) effectively amounts to a redefinition of our vacuum state. To fix a vacuum state, we need to impose conditions that uniquely define the mode functions $v_k(\eta)$ and $v_k^*(\eta)$, for example, by defining boundary conditions for them.

For the problem at hand, a quantum field during inflation, there is a standard notion of vacuum, and thus a standard choice of initial conditions. To motivate the choice, consider as a simplification the case of a perfect de Sitter expansion, with a constant Hubble rate, $H = \text{const}$, and an exact relationship between conformal time, η , and the comoving Hubble scale, $\eta = -1/(aH)$, with $\eta \in (-\infty, 0)$. In this case, Eq. (1.54), simplifies to

$$v_k'' + \left[k^2 - \left(2 - \frac{m^2}{H^2} \right) \frac{1}{\eta^2} \right] v_k = 0, \quad (1.59)$$

which approaches a simple harmonic oscillator in the limit $-k\eta \gg 1$:

$$v_k'' + k^2 v_k = 0 \quad \text{for} \quad -k\eta \rightarrow \infty. \quad (1.60)$$

In this limit, Eq. (1.59) admits the standard choice of mode functions we would use in flat spacetime, namely a positive-frequency plane-wave solution,

$$v_k(\eta) \rightarrow \frac{1}{\sqrt{2k}} e^{-ik\eta} \quad \text{for} \quad -k\eta \rightarrow \infty. \quad (1.61)$$

This choice of boundary condition uniquely determines the mode functions v_k , and thus also corresponds to a unique selection of a vacuum state. Boundary conditions of this type are called the *Bunch–Davies initial conditions*, and the associated vacuum is the *Bunch–Davies vacuum* [185]. It corresponds to the vacuum state that we can define in the far distant past, $\eta \rightarrow -\infty$, for a comoving observer, when the Hubble radius, H^{-1} , is far greater than the reduced wavelength $\lambda = a/k \rightarrow 0$ for any quantum fluctuation of comoving momentum $\mathbf{k}$.¹⁰ In a flat FLRW spacetime, the Hubble radius H^{-1} gives an estimate of local curvature such that a mode with $k \gg aH$ can effectively be viewed as oscillating in a locally flat spacetime. Since we are working in the Heisenberg picture, time dependence is carried by the operators χ and π , and the Bunch–Davies vacuum is time-independent. It is this vacuum which we have implicitly assumed in section 1.2.1, when we identified the inflaton background field as a vacuum expectation value, e.g., $\varphi = \langle \phi \rangle$.

1.3.2 Super-horizon evolution and classicalization

Having fixed the initial conditions for the mode functions v_k , one can now solve the mode equation in Eq. (1.54) to obtain their time evolution. To keep things simple, we will continue to assume

¹⁰For simplicity, we define the reduced wavelength as $\lambda = \lambda/(2\pi)$ with λ the physical wavelength.

perfect de Sitter space. Assuming $m^2 < 3H^2/2$, a solution to Eq. (1.59) with the correct boundary conditions (up to a phase) is then given by

$$v_k(\eta) = \frac{\sqrt{\pi|\eta|}}{2} H_\nu^{(1)}(-k\eta) \quad \text{with} \quad \nu = \sqrt{\frac{9}{4} - \frac{m^2}{H^2}}. \quad (1.62)$$

where $H_\nu^{(1)}(x)$ is the Hankel function of the first kind (see, e.g., Ref. [186]). Using the asymptotics of the Hankel functions, we can now consider also the late-time limit, $-\eta k \ll 1$. We get,

$$v_k(\eta) \rightarrow -i \frac{1}{\sqrt{\pi} k^\nu 2^{1-\nu}} \Gamma(\nu) |\eta|^{1/2-\nu} \quad \text{for} \quad -k\eta \rightarrow 0, \quad (1.63)$$

where Γ denotes the Euler Gamma function. For $m^2 \ll H^2$, this corresponds to a linear growth with the scale factor $a(\eta)$. Besides the fact that this is a growing solution, the more important observation is that the mode function v_k has stopped oscillating in the late-time limit. To obtain physical intuition of this behavior, remember that $-\eta k \ll 1$ implies $k \ll aH$, which corresponds to the mode function's reduced physical wavelength, $\lambda = a/k$ having been stretched to a size much larger than the Hubble radius, H^{-1} . We say that the mode has *frozen* on super-horizon scales.

Another important implication of Eq. (1.63) can be understood by considering the original field operators χ and π in the limit $\eta \rightarrow 0$. From Eq. (1.53), one finds

$$\chi(\eta, \mathbf{x}) \propto |\eta|^{1/2-\nu} \int \frac{d^3 \mathbf{k}}{(2\pi)^{3/2}} \frac{1}{k^\nu} \left(\hat{a}_{\mathbf{k}} - \hat{a}_{-\mathbf{k}}^\dagger \right) e^{i\mathbf{k} \cdot \mathbf{x}}, \quad (1.64a)$$

$$\pi(\eta, \mathbf{x}) \propto |\eta|^{-1/2-\nu} \int \frac{d^3 \mathbf{k}}{(2\pi)^{3/2}} \frac{1}{k^\nu} \left(\hat{a}_{\mathbf{k}} - \hat{a}_{-\mathbf{k}}^\dagger \right) e^{i\mathbf{k} \cdot \mathbf{x}} \propto -\frac{1}{\eta} \chi(\eta, \mathbf{x}). \quad (1.64b)$$

The direct proportionality between χ and π means that the two fields commute for $\eta \rightarrow 0$, and their equal-time commutator vanishes. The standard interpretation of this is that the quantum fields have classicalized due to the rapid expansion of spacetime. Therefore, we view a quantum fluctuation with wavenumber k as classical once it reaches super-horizon scales, $k \ll aH$ [187].

As a last step, we should remember that the physical field is not χ , but rather $\phi = \chi/a$. For a light scalar field, $m^2 \ll H^2$, we have seen that the mode function for χ grows linearly with the scale factor in the late time limit. This implies that modes of the field ϕ in fact become constant when their wavelength has grown to super-horizon size. In particular, this has important implications for inflationary perturbations, as we shall discuss in Sec. 1.5.

The above conclusions indicate remarkable features about quantum fields in de Sitter space. However, during inflation, we do not in actuality have exponential expansion, and the Hubble rate will mildly drift over time. However, these effects are typically small (in particular because $\varepsilon_V \ll 1$ during slow-roll inflation) and we can therefore assume that our conclusions also apply to a period of inflation, and in particular to the inflaton field itself.

1.3.3 The spectrum of a light scalar field during inflation

As a last point, we now want to determine the spectrum of a light scalar field ϕ in de Sitter space, which establishes an important result for the next section.

Using Eq. (1.53) and $\phi = \chi/a$, we first determine the two-point correlation function of ϕ , which is given by

$$\langle \phi^2 \rangle = \frac{1}{a^2} \int_0^\infty \frac{dk}{k} \frac{k^3}{2\pi^2} |v_k(\eta)|^2. \quad (1.65)$$

Remembering the asymptotic behavior of the mode function v_k for $k \rightarrow \infty$, we see that this integral diverges in the UV. This should not come as a great surprise to us, considering that any mode $v_k(\eta)$ with $-k\eta \ll 1$ effectively corresponds to a quantum harmonic oscillator in its vacuum configuration (see Eq. (1.61)), and we are integrating over infinitely many of these. This divergence is then just one of the usual UV divergences of QFT. A simple method to regularize it is to introduce an effective cut-off at the horizon scale, $k_H = aH$. Since modes with $k \ll aH$ have already stopped oscillating, and we view them as classical, this effectively removes any quantum oscillations in Eq. (1.65).¹¹ Keeping only leading terms in aH/k , we then get

$$\langle \phi^2 \rangle_{\text{reg.}} = \left(\frac{H}{2\pi} \right)^2 \frac{2^{2\nu} \Gamma^2(\nu)}{2\pi} \int_0^{k_H} \frac{dk}{k} \left(\frac{k}{aH} \right)^{3-2\nu} = \left(\frac{H}{2\pi} \right)^2 \frac{2^{2\nu} \Gamma^2(\nu)}{2\pi(3-2\nu)}. \quad (1.66)$$

From this result, one may then infer the spectrum of scalar field perturbations as a function of wavenumber by taking its logarithmic derivative with respect to k . This yields

$$\Delta\phi_k \equiv \sqrt{\frac{d\langle \phi^2 \rangle}{d \log k}} = \left(\frac{H}{2\pi} \right) \frac{2^\nu \Gamma(\nu)}{\sqrt{2\pi}} \left(\frac{k}{aH} \right)^{3/2-\nu} \rightarrow \left(\frac{H}{2\pi} \right) \quad \text{for} \quad \frac{m^2}{H^2} \rightarrow 0 \quad (1.67)$$

This is the well-known, and fundamental result that the spectrum of a light scalar field is nearly scale-invariant for modes that have excited the horizon, i.e., with $k \ll aH$. Not only is the spectrum nearly flat, it is also time independent as a consequence of the freezing of super-horizon modes for nearly massless scalar fields. Moreover, this spectrum is Gaussian: If we had computed higher-order correlation functions of ϕ , Eq. (1.53) would imply that odd-numbered n -point correlators will vanish, while even-numbered ones are given in terms of two-point correlators.

1.4 Act II: Primordial perturbations from slow-roll inflation

In the previous section, we have seen that vacuum fluctuations of massless scalar fields are continuously stretched to super-horizon scales during inflation, where they then classicalize and freeze. In slow-roll inflation, this process takes place for the perturbations in the inflaton field and the spacetime metric. This sources a spectrum of nearly scale-invariant and Gaussian primordial perturbations, which then manifest themselves in later epochs of cosmic expansion as scalar perturbations to the Universe's homogeneous density [22–26], and as primordial gravitational waves (GWs) [27–31].

In this section, we briefly establish these predictions for slow-roll inflation, before reviewing in the next section how measuring the traces of these primordial perturbations in the CMB or with GW observatories can allow insights into the physics of inflation. To this end, we first give a brief overview of important ideas that we need to describe perturbations to a flat FLRW spacetime before we present the standard results for inflationary perturbations from slow-roll inflation.

It should be noted that part III of this thesis is fully dedicated to the discussion of inflationary perturbations, where we consider how the results we state below change when we account for gauge-field production in the model of axion inflation. As part of this later discussion, we naturally derive many of the results that appear in what follows, and we therefore give only a concise description of this topic here. The discussion of inflationary perturbations can also be found in many textbooks and lecture notes on inflation. We mainly rely on Refs. [137, 138, 190] as the basis for our discussion, with additional insights taken from Refs. [191, 192].

¹¹There are more sophisticated techniques for regularizing Eq. (1.65), for example Pauli–Villars regularization or adiabatic subtraction [188, 189], however we will not use these in this thesis.

1.4.1 Preliminary considerations

Thus far in this chapter, we have always assumed a purely homogeneous and isotropic Universe described by the flat FLRW metric in Eq. (1.23). Now, we want to consider perturbations to this homogeneous and isotropic background as these are what ultimately generates the primordial density perturbations and gravitational waves during inflation. In slow-roll inflation, these are given by perturbations to the inflaton field, but also to the spacetime metric, as any inhomogeneities in the distribution of matter are necessarily reflected in the metric according to the Einstein field equations (1.3). The mathematical description of such perturbations is given by the formalism of cosmological perturbation theory which one can use to study how perturbations to FLRW spacetime evolve in time. More details on cosmological perturbation theory will be given in chapter 7, and we here only summarize important concepts that we need in what follows.

Linear perturbations in slow-roll inflation can be described by a perturbative Ansatz for the metric and the inflaton field

$$g_{\mu\nu}(\eta, \mathbf{x}) = a(\eta)^2 (\eta_{\mu\nu} + \delta h_{\mu\nu}(\eta, \mathbf{x})) \quad \text{and} \quad \phi(\eta, \mathbf{x}) = \varphi(\eta) + \delta\phi(\eta, \mathbf{x}). \quad (1.68)$$

Here, $\eta_{\mu\nu}$ is the Minkowski metric, a is the scale factor and φ is the background inflaton field. Their evolution is given by the Friedmann and the Klein–Gordon equations in Eq. (1.26), while $\delta h_{\mu\nu}$ and $\delta\phi$ are treated as small, inhomogeneous perturbations to FLRW space. The evolution of these perturbations can be derived by expanding the Einstein field equations (1.3), the energy–momentum tensor of ϕ in Eq. (1.21), and the equations of motion for a scalar field in curved spacetime in Eq. (1.22) to linear order in $\delta h_{\mu\nu}$ and $\delta\phi$.

The perturbations variables $\delta h_{\mu\nu}$ and $\delta\phi$ make up the spectrum of inflationary perturbations. In cosmological perturbation theory, these perturbations are best described as spatial three-tensors, as this takes advantage of the fact that the background FLRW spacetime on which they evolve is invariant under spatial translations and spatial rotations, but not under Lorentz boosts or time translations. We denote three-tensors by Latin indices to distinguish them from four-tensors, and we define them such that indices are raised and lowered by the Euclidean metric, δ_{ij} , i.e., $X^i_j = \delta^{ik} X_{kj}$. It is then advantageous to describe perturbations in Fourier space. We define the Fourier transform of any perturbation $X(\eta, \mathbf{x})$ (e.g., $\delta\phi$) as

$$X_{\mathbf{k}}(\eta) \equiv \int \frac{d^3\mathbf{k}}{(2\pi)^{3/2}} X(\eta, \mathbf{x}) e^{-i\mathbf{k}\cdot\mathbf{x}}. \quad (1.69)$$

The rotational invariance of FLRW spacetime also entails that the spacetime perturbations $\delta h_{\mu\nu}$ and $\delta\phi$ are best decomposed into a set of scalars, X , transverse vectors X_i with $k^i X_i = 0$, and transverse traceless tensors X_{ij} with $k^i X_{ij} = X^i_i = 0$. At linear order in perturbation theory, these scalar, vector, and tensor perturbations are decoupled, and they evolve independently of one another. In particular, the spectrum of inflationary perturbations is described by scalar perturbations in the spacetime metric $\delta h_{\mu\nu}$ and inflaton perturbations $\delta\phi$, which collectively can be understood as creating primordial density perturbations, as well as tensor perturbations in $\delta h_{\mu\nu}$ that can be understood as primordial gravitational waves, similarly to flat Minkowski space. Vector perturbations prove to be irrelevant, as they generically decay with cosmic expansion. More details on this scalar–vector–tensor decomposition are given in chapter 7.

In what follows, we will therefore discuss the spectra of inflationary scalar and tensor perturbations. When describing these, it is important to note that inflationary perturbations that are created during inflation are inherently stochastic in nature, as they originate from quantum fluctuations. It is common practice in cosmology to phrase results for stochastic quantities in terms of *power spectra*. Let X be a random variable, we define its power spectrum, $\mathcal{P}_X(\eta, k)$, via the two-point

correlator,

$$\langle X_{\mathbf{k}}(\eta) X_{\mathbf{k}}(\eta) \rangle = \delta^{(3)}(\mathbf{k} + \mathbf{k}') \frac{2\pi^2}{k^3} \mathcal{P}_X(\eta, k), \quad (1.70)$$

where the implicit definition of $\mathcal{P}_X(\eta, k)$ reflects the symmetry of the FLRW background: translational invariance enforces momentum conservation, $\delta^{(3)}(\mathbf{k} + \mathbf{k}')$, while rotational invariance implies $\mathcal{P}_X(\eta, k)$ only depends on $k = |\mathbf{k}|$. Assuming furthermore that X is Gaussian distributed, all relevant statistical information on its spectrum is then contained in $\mathcal{P}_X(\eta, k)$.¹²

1.4.2 Primordial scalar perturbations

To discuss primordial scalar perturbations, we first need to introduce the notion of gauge invariance in perturbation theory, which is a manifestation of the freedom to choose coordinates in GR. This freedom implies that, e.g., the metric perturbation, $\delta h_{\mu\nu}$ in Eq. (1.68) is only defined up to an infinitesimal coordinate transformation of the kind

$$x^\mu \rightarrow x^\mu + \xi^\mu(\eta, \mathbf{x}), \quad (1.71)$$

with the shift ξ^μ acting as a perturbation, $|\xi^\mu| \ll 1$. Thus, as in electrodynamics, we have a gauge freedom in choosing ξ^μ which can be used to simplify calculations, but it also entails that physical results should ultimately be expressed in terms of gauge-invariant quantities.

In particular, ξ^μ contains two scalar components that can be chosen to remove two of the scalar metric perturbations or the inflaton perturbation $\delta\phi$ from the mathematical description of scalar perturbations. For instance, $\delta\phi$ can be removed by an appropriate choice of ξ^0 , as we can note by considering the change in the scalar field under an infinitesimal time translation

$$\varphi(\eta) + \delta\phi(\eta, \mathbf{x}) \rightarrow \varphi(\eta) + \delta\phi(\eta, \mathbf{x}) + \varphi'(\eta)\xi^0(\eta, \mathbf{x}) \quad \text{for } \eta \rightarrow \eta + \xi^0(\eta, \mathbf{x}). \quad (1.72)$$

The new inhomogeneous piece, $\varphi'(\eta)\xi^0$, constitutes a shift to $\delta\phi$, which we may even choose such that $\delta\phi$ vanishes entirely. However, this does not mean that the physical effect of $\delta\phi$ is unimportant, as such a shift simply moves it to perturbations of the spacetime metric.

In our thesis, we will use this gauge freedom to work in *Newtonian gauge* (or *longitudinal gauge*), which removes two scalar perturbations from the spacetime metric. The perturbed FLRW metric in this gauge can be written as the line element¹³

$$ds^2 = a^2(\eta) \left[(1 + 2\Psi(\eta, \mathbf{x})) d\eta^2 - (1 - 2\Phi(\eta, \mathbf{x})) \delta_{ij} dx^i dx^j \right], \quad (1.73)$$

where Ψ and Φ are the gauge-invariant *Bardeen potentials*, which one can interpret as generalizations of the usual Newtonian potential, hence the name of our gauge.

The spectrum of scalar perturbations from inflation is described by the gauge-invariant *comoving curvature perturbation*, which in our case is given by

$$\zeta = \Phi + \frac{H}{\dot{\phi}} \delta\phi. \quad (1.74)$$

¹²If X encompasses multiple degrees of freedom, e.g., polarization indices, our definition implicitly assumes they are summed over. For example, the power spectrum of a tensor perturbation, X_{ij} , is

$$\langle X_{\mathbf{k}}^{ij}(\eta) X_{ij, \mathbf{k}}(\eta) \rangle = \delta^{(3)}(\mathbf{k} + \mathbf{k}') \frac{2\pi^2}{k^3} \mathcal{P}_X(\eta, k).$$

¹³Here, we only consider scalar perturbations to the spacetime metric, as they are decoupled from vector and tensor perturbations, see above.

The name of this variable reflects the fact that it is directly related to the perturbations of the Ricci scalar for spatial hypersurfaces when working in comoving coordinates.

To determine the time evolution of ζ , one needs to find its equation of motion. This is done by deriving a set of linearized Einstein equations for scalar perturbations (see Eq. (7.26)), and combining it with the Klein–Gordon equation from Eq. (1.22) for the metric in Eq. (1.73). One then finds the following evolution equation in Fourier space;

$$\zeta_{\mathbf{k}}'' + 2\frac{z'}{z}\zeta_{\mathbf{k}}' + k^2\zeta_{\mathbf{k}} = 0 \quad \text{with} \quad z = \frac{a\dot{\phi}}{H}. \quad (1.75)$$

By a change of variable, this can be written as

$$v_{\mathbf{k}}'' + \left(k^2 - \frac{z''}{z}\right)v_{\mathbf{k}} = 0 \quad \text{with} \quad v_{\mathbf{k}} \equiv z\zeta_{\mathbf{k}}, \quad (1.76)$$

where $v_{\mathbf{k}}$ is the famous *Mukhanov–Sasaki variable* [193, 194]. Note that in the limit of perfect de Sitter space, $\varepsilon_V \rightarrow 0$, we have that $z''/z \rightarrow a''/a$, and Eq. (1.76) is just the mode equation for a massless scalar field, like in Eq. (1.54).¹⁴ Having already discussed this scenario at length, the next steps are clear: we quantize $v_{\mathbf{k}}$, introduce Bunch–Davies initial conditions, and determine its mode functions by solving Eq. (1.76). The final result for $\zeta_{\mathbf{k}} = v_{\mathbf{k}}/z$ is then qualitatively the same as for a massless scalar field; it is Gaussian, frozen on super horizon scales, and has a nearly scale-invariant spectrum.

Phrasing this final result in terms of the power spectrum for ζ , one obtains

$$\mathcal{P}_\zeta(\eta_k, k) = \left(\frac{H^2}{2\pi|\dot{\phi}|}\right)_{\eta_k}^2, \quad (1.77)$$

where η_k denotes the time when $\zeta_{\mathbf{k}}$ exits the horizon, i.e., $k = a(\eta_k)H(\eta_k)$ and the subscript “ η_k ” on the right-hand side denotes that the quantities are evaluated at this time. For later times during inflation, $\eta > \eta_k$, the value of this power spectrum will be frozen at this amplitude.

A useful parametrization of Eq. (1.77) is

$$\mathcal{P}_\zeta(\eta_k, k) = A_s \left(\frac{k}{k_*}\right)^{n_s-1}, \quad (1.78)$$

where A_s is the *amplitude of scalar perturbations*, n_s is the *scalar spectral index* and k_* is a reference scale at which A_s and n_s are defined. From Eq. (1.77), we then find the standard prediction for slow-roll inflation,

$$A_s = \mathcal{P}_\zeta(\eta_*, k_*) = \left(\frac{H^2}{2\pi|\dot{\phi}|}\right)_{\eta_*}^2 = \frac{1}{24\pi^2\varepsilon_{V,*}} \frac{V_*}{M_{\text{P}}^4}, \quad (1.79a)$$

$$n_s - 1 = \frac{d \log \mathcal{P}_\zeta}{d \log k} \Big|_{k_*} \simeq \frac{1}{k} \frac{d \log \mathcal{P}_\zeta}{d \eta} \Big|_{\eta_*} \simeq -6\varepsilon_{V,*} + 2\eta_{V,*}, \quad (1.79b)$$

where the subscript “ $*$ ” indicates that the quantity is to be evaluated at the time when k_* exits the horizon, $\eta_* \simeq -1/(a_*H_*) = -1/k_*$, and we have kept terms to linear order in the slow roll

¹⁴ Another approach to obtain this equation is to directly expand the action Eq. (1.20) to second order in perturbation theory, and to use it to define an action for the Mukhanov–Sasaki variable. This is usually done in either *comoving gauge* ($\delta\phi = 0$) or *spatially-flat gauge* ($\delta\phi = (\dot{\phi}/H)\zeta$), where the second-order perturbed action is particularly simple.

parameters ε_V and η_V . For slow-roll inflation, $\varepsilon_V, |\eta_V| \ll 1$, and the above prediction is indeed that of a nearly scale-invariant curvature power spectrum.

Before we move on to discuss tensor perturbations, we should mention another important feature of the scalar perturbations induced by slow-roll inflation, namely that they are *adiabatic perturbations*. To understand this, consider the implications of Eq. (1.72). We have seen that an infinitesimal shift in time, $\eta \rightarrow \eta + \delta\eta$, is equivalent to a perturbation of the inflaton field, i.e., $\delta\phi(\eta, \mathbf{x}) = -\varphi'(\eta)\delta\eta(\eta, \mathbf{x})$. Interpreting this in a functional sense, the evolution of $\delta\phi(\eta, \mathbf{x})$ at every point in space is entirely determined by the background field $\varphi(\eta)$. When scalar perturbations induced by inflation reenter the Hubble horizon during RD, this means that all sub-components of the thermal bath, i.e., dark matter, baryonic matter, photons, etc., are affected by a common perturbation to their density, and they all evolve depending purely on the evolution of the background energy density in radiation. Thus, we say that slow-roll inflation induces adiabatic perturbations, as opposed to *entropy perturbations* (or *isocurvature perturbations*). The curvature perturbation variable ζ in Eq. (1.74) has the useful property that it is conserved on super-horizon scales for adiabatic perturbations, which is why we recover these properties in the case at hand.

1.4.3 Primordial tensor perturbations

The study of tensor perturbations proves to be much simpler than that of scalar perturbations. There is no gauge freedom for tensor perturbations, and $\delta h_{\mu\nu}$ contains only a single, gauge-invariant tensor component corresponding to two propagating degrees of freedom. The perturbed FLRW metric in the presence of tensor perturbations is

$$ds^2 = a^2 [d\eta^2 - (\delta_{ij} + h_{ij}(\eta, \mathbf{x}))dx^i dx^j] , \quad (1.80)$$

where we interpret h_{ij} as gravitational waves, as we would in Minkowski space.

For slow-roll inflation, there is no source for tensor perturbations at linear order, and the evolution equation for h_{ij} simply reads,

$$h_{ij}'' + 2(aH)h_{ij}' - \Delta h_{ij} = 0 . \quad (1.81)$$

Again, it is in the form of a Klein–Gordon equation in an expanding spacetime, so our usual strategy of quantization applies. The conclusion is then the same as for scalar perturbations: tensor perturbations from slow-roll inflation are Gaussian, frozen on super-horizon scales, and their spectrum is nearly scale-invariant.

On horizon exit, the tensor power spectrum predicted by slow-roll inflation is then given by

$$\mathcal{P}_h(\eta_k, k) = \left(\frac{2}{\pi^2} \frac{H^2}{M_{\text{P}}^2} \right)_{\eta_k} . \quad (1.82)$$

with η_k defined as in Eq. (1.79a). This result is conventionally parametrized as

$$\mathcal{P}_h(\eta_k, k) = r A_s \left(\frac{k}{k_*} \right)^{n_t} , \quad (1.83)$$

where $r = \mathcal{P}_h(k_*)/\mathcal{P}_\zeta(k_*)$ is the *tensor-to-scalar ratio* and n_t is the *tensor spectral index*. The predictions for slow-roll inflation are

$$r = \frac{\mathcal{P}_h(\eta_*, k_*)}{\mathcal{P}_\zeta(\eta_*, k_*)} = \left(8 \frac{\dot{\varphi}^2}{H^2 M_{\text{P}}^2} \right)_{\eta_*} \simeq 16\varepsilon_{V,*} , \quad (1.84a)$$

$$n_t = \left. \frac{d \log \mathcal{P}_h}{d \log k} \right|_{k_*} = \frac{1}{k} \left. \frac{d \log \mathcal{P}_h}{d \eta} \right|_{\eta_*} \simeq -2\epsilon_{V,*}. \quad (1.84b)$$

Notably, slow-roll inflation predicts a proportionality between the parameters n_t and r . Since ϵ_V is positive, $\mathcal{P}_h$ is *red-tilted* for slow-roll inflation, i.e., the power spectrum grows when going to smaller k (as opposed to *blue-tilted*; the opposite case).

1.5 Act III: The observable fingerprint of inflation

Now that we have seen the basic predictions for the scalar and tensor power spectra from slow-roll inflation, we can close this chapter by discussing how they can manifest themselves in our present-day Universe, and what the current experimental constraints on the dynamics of slow-roll inflation can be inferred from these predictions. To this end, we first consider how inflationary perturbations of different wavenumber k manifest themselves on different physical length scales today. Then, we review the current constraints on inflation from modern measurements of the CMB's polarization and temperature anisotropies. Finally, we discuss other direct and indirect observable signatures that inflationary perturbation can leave beyond the CMB.

1.5.1 Super-horizon conservation: wavenumber as a clock variable

When discussing observable constraints on inflation, an important concept is that inflationary perturbations which are generated at earlier times during inflation manifest themselves on larger physical distance scales than those that are generated at later times during inflation. As we shall elaborate upon in what follows, this fundamentally builds on the fact that the amplitude of the scalar and tensor power spectrum for a given wavenumber k freezes upon horizon exit ($k \simeq aH$), and that the physical distance scales corresponding to the wavenumber k are continuously stretched to larger scales by cosmic expansion.

Consider a perturbation $X_{\mathbf{k}}$ (e.g., ζ , or h_{ij}) with its comoving wavenumber $k = |\mathbf{k}|$ corresponding to a reduced physical wavelength $\lambda_{\text{phys},k}(t) = a(t)/k$ at time t . Because the comoving Hubble scale increases during inflation, the perturbation exits the horizon when $k = a_{\text{out},k}H_{\text{out},k}$, where we implicitly define the time of horizon exit for the wavenumber k as $t_{\text{out},k}$. At this point, the amplitude of its power spectrum, $\mathcal{P}_X(t_{\text{out},k}, k)$ will be frozen. This corresponds to the physical wavelength of $X_{\mathbf{k}}$ being of the size as the Hubble horizon at $t_{\text{out},k}$, i.e., $\lambda_{\text{phys},k}(t_{\text{out},k}) = H_{\text{out},k}^{-1}$. The amplitude of $\mathcal{P}_X$ then remains frozen for the remaining duration of inflation. However, once inflation ends, accelerated expansion stops, and aH begins to decrease, as we have explained in Sec. 1.1. Therefore, the wavenumber k reenters the horizon once $k = a_{\text{in},k}H_{\text{in},k}$, at a time $t_{\text{in},k} > t_{\text{out},k}$. When this happens, $X_{\mathbf{k}}$ acts as a classical stochastic perturbation to the homogeneous Universe with a reduced wavelength $\lambda_{\text{phys},k}(t_{\text{in},k}) = a(t_{\text{in},k})/k$ with its amplitude being determined by the conserved value of its power spectrum, $\mathcal{P}_X(t_{\text{out},k}, k)$.

If we now consider this chronology for perturbations of two distinct wavenumbers, $k_1 > k_2$, we find that k_2 exits the horizon during inflation before k_1 does, while the horizon reentry for k_1 occurs before k_2 . Comparing their respective reduced physical wavelengths, $\lambda_{\text{phys},i}(t) = 2\pi a(t)/k_i$ at any time t , we see that k_2 corresponds to larger wavelengths than k_1 does. Thus, large-scale perturbations originate from early times during inflation, while small scale perturbations are sourced at later times. An illustration of this idea is given in Fig. 1.4.

In this way, by measuring the spectrum of perturbations at a reduced physical wavelength $\lambda_{\text{phys}} = a_0/k$ today, we can use our knowledge of the cosmic evolution to infer the time of its horizon exit, when $k = a_{\text{out},k}H_{\text{out},k}$ during inflation. This can be done up to the knowledge of

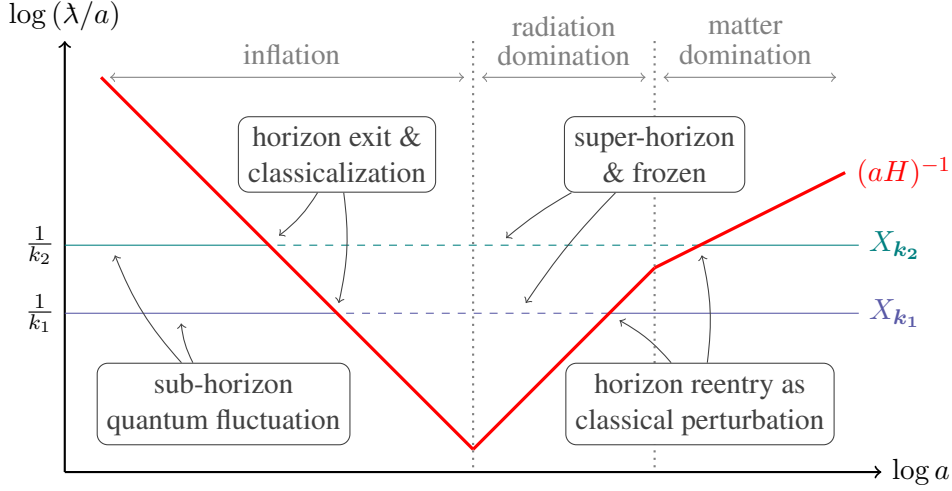


FIGURE 1.4: A sketch of two inflationary perturbation modes X_{k_1} (blue) and X_{k_2} (teal) with wavenumbers $k_1 > k_2$ exiting the Hubble horizon (red) during inflation and reentering it during radiation domination (RD) or matter domination (MD). During inflation, the mode of larger wavelength (X_{k_2}) exits the horizon before the one of smaller wavelength (X_{k_1}) does. After inflation, the modes are frozen (indicated by dashed lines) until they reenter as classical perturbations during RD (X_{k_1}) and MD (X_{k_2}). Note the reversed order of horizon reentry with respect to horizon exit.

reheating, whose dynamics are model-dependent (see Sec. 1.2.3). Assuming instantaneous reheating, one can determine the following useful relationship between the time when a wavenumber k has exited the horizon before the end of inflation, and its physical wavenumber today:

$$\frac{k}{a_0} \simeq 2.3 \times 10^{25} \left(\frac{H_{\text{end}}}{M_{\text{P}}} \right)^{1/2} e^{-N_k} \text{ Mpc}^{-1}. \quad (\text{inst. reheating}) \quad (1.85)$$

Here, $N_k = \log(a_{\text{end}}/a_{\text{out},k})$ denotes the number of e -folds before the end of inflation at which k has exited the horizon, i.e., $k = a_{\text{out},k} H_{\text{out},k}$. This expression can be derived by assuming that the Hubble rate during inflation varies slowly, $H_{\text{out},k} \simeq H_{\text{end}}$, using Eq. (1.11) to relate temperature T to the scale factor a , and then using Eq. (1.48) to relate H_{end} to the reheating temperature.

Equation (1.85) is useful to establish a one-to-one correspondence between a perturbation of wavelength $2\pi a_0/k$ today, and the time period during inflation at which it has left the horizon, allowing us to essentially use the wavenumber k as a clock variable during inflation. In this sense, experiments that probe physical distances corresponding to physical wavenumbers k/a_0 today, probe the physics of inflation approximately at a time N_k as given by Eq. (1.85).

1.5.2 Slow-roll inflation and the anisotropies in the CMB

At present, the only direct observable constraints on inflation come from the measurements of the polarization and temperature anisotropies of the CMB, most prominently those by the Planck satellite [14, 16]. These measurements constrain the scalar and tensor power spectra in Eqs. (1.78) and (1.83) from slow-roll inflation at physical wavenumbers that are thought to have left the inflationary horizon at around 50 – 60 e -folds before the end of inflation, depending on the details of reheating. In what follows, we summarize current experimental constraints on the scalar and tensor power spectrum from inflation, and then a discussion of what can be inferred from these results. For reviews on CMB physics, see e.g., Refs. [195, 196].

Constraints on scalar and tensor perturbations

Modern CMB measurements are accurate to a degree where they can be used to infer both the amplitude A_s and the spectral tilt n_s of the primordial curvature power spectrum in Eq. (1.78). These values are primarily determined through measurements of the CMB's temperature anisotropies (although polarization anisotropies provide complementary information, see below). During radiation domination, primordial curvature perturbations can be understood as perturbations to the homogeneous background density of radiation, ρ_r . Such perturbations induce temperature anisotropies in this background photon fluid (roughly, because $\rho_r \propto T^4$). Thus, the amplitude of the curvature power spectrum A_s directly sets the amplitude of temperature anisotropies, while n_s quantifies the relative amount of temperature anisotropies on small scales versus large scales.

The last dataset provided by the Planck satellite [14, 16] (PR3) finds that A_s and n_s are consistent with a nearly scale-invariant primordial curvature power spectrum,

$$\log(10^{10} A_s) = 3.044 \pm 0.014, \quad n_s = 0.9649 \pm 0.0042. \quad (\text{at } 68\% \text{ C.L.})$$

These quantities are inferred at a pivot scale $k_{\text{CMB}} = 0.05 \text{ Mpc}^{-1}$. The final Planck data has been recently reassessed in a fourth data release [197] (PR4), and gives consistent results with PR3 [198]. Recently, ground-based CMB observatories have provided a complementary measurement to Planck: the Atacama Cosmology Telescope (ACT) [33, 199], and the South Pole Telescope (SPT) [32]. Overall these measurements are in agreement with Planck. It has been noted, however, that a combined analysis of ACT and Planck results alongside measurements of baryon acoustic oscillations using the second data release of the Dark Energy Spectroscopic Instrument (DESI) [200] results in a slightly higher value for the spectral tilt, $n_s = 0.9752 \pm 0.0030$, which is in 2σ tension with the PR3 result [199]. Both PR3 and ACT also test for a running of the spectral index, $dn_s/d \ln k$, but find results consistent with zero [14, 33].

Primordial tensor perturbations also affect CMB photons, which is most prominently detectable by measuring the polarization anisotropies in the CMB. Such anisotropies arise due to temperature anisotropies at the time of last scattering and through gravitational waves affecting photon polarizations as they traverse the Universe after the CMB's emission. The resulting statistical anisotropies are classified in terms of E- and B-mode polarization, corresponding to parity-even (E) and parity-odd (B) contributions. Primordial scalar perturbations only induce E-mode polarization, while tensor perturbations induce both E- and B-mode polarization. Therefore, measuring the B-mode power spectrum is the main target for determining the amplitude and spectral tilt of the tensor power spectrum, as given by r and n_t in Eq. (1.83). However, this measurement is limited due to B-mode polarization from gravitational lensing and galactic foreground emissions, thus CMB measurements only place upper limits on the tensor-to-scalar ratio r and leave n_t unconstrained.

The BICEP/Keck collaboration places an upper limit $r < 0.036$ at 95% C.L. [201] (BK18) when combining their data with PR3. This bound has been re-evaluated using PR4, yielding the improved upper bound [202]¹⁵

$$r < 0.032. \quad (\text{at } 95\% \text{ C.L.})$$

SPT also constrains the tensor-to-scalar ratio, finding $r < 0.25$ [32], however we rely on the more constraining PR4+BK18 limit in this thesis. Tensor perturbations at CMB scales may be detectable by future CMB missions like CMB-S4 [204].

¹⁵See also Ref. [203] which find $r < 0.03$ and $-1.35 < n_t < 3.40$ at 95% C.L. when relaxing the assumption that r and n_t are proportional as predicted by slow-roll inflation (see Eq. (1.84)).

Implications for slow-roll inflation

Combining the upper limit on r and the measurement of A_s the amplitude in tensor perturbations at CMB scales is constrained as $rA_s < 6.3 \times 10^{-11}$. Taking the prediction for the amplitude of tensor perturbations in Eq. (1.82), this can be translated into an upper limit on the scale of inflation,

$$H_{\text{infl.}} < 1.8 \times 10^{-5} M_{\text{P}} \quad \text{or} \quad H_{\text{infl.}} < 4.3 \times 10^{13} \text{ GeV}. \quad (1.86)$$

This upper limit can also be rewritten as a constraint on the amplitude of the scalar potential at CMB scales,

$$V(\phi_{\text{CMB}}) \lesssim 9.3 \times 10^{-10} M_{\text{P}}^4. \quad (1.87)$$

Furthermore, assuming the maximal reheating temperature is obtained for instantaneous reheating, one finds

$$T_{\text{reh}} \lesssim 5.6 \times 10^{15} \text{ GeV}. \quad (1.88)$$

Note that we used the value $T_{\text{reh}} \simeq 5.6 \times 10^{15} \text{ GeV}$ in Eq. (1.18) to deduce that $N_{\text{infl.}} \gtrsim 60$ e -folds of inflation are needed to solve the flatness problem.

Using Eq. (1.85), one can trace back the scales constrained by CMB measurements to the time of their horizon exit during inflation. Taking the pivot scale $k_{\text{CMB}} = 0.05 \text{ Mpc}^{-1}$ as exemplary one finds that CMB measurements constrain the physics of inflation at

$$N_{\text{CMB}}^{\text{inst.}} \simeq \log \left(\frac{2.3 \times 10^{25} \text{ Mpc}^{-1}}{k_{\text{CMB}}} \right) + \frac{1}{2} \left(\frac{H_{\text{infl.}}}{M_{\text{P}}} \right) \lesssim 55.9, \quad (1.89)$$

where the last inequality comes from Eq. (1.86). Thus, even when disregarding the horizon and flatness problems, if the aim of inflation is purely to explain the temperature and polarization anisotropies in the CMB, current data requires about 60 e -folds of accelerated expansion at a high scale of inflation. Additionally, as the CMB probes the power spectra for $k \sim 10^{-3} - 10^{-1} \text{ Mpc}^{-1}$, using Eq. (1.85) we find a correspondence to a range of $\Delta N \simeq 5$ e -folds during inflation.

Using the relationship between the slow-roll parameters ε_V and η_V in Eq. (1.30), and their relationship to A_s , n_s and r in Eqs. (1.79) and (1.84), CMB data can further be used to constrain a large set of established slow-roll potentials. These constraints are summarized in overview figures as in Fig. 1.5, which we have taken from Ref. [14] and which shows the results for Planck PR3 data in combination with the 2015 data from the BICEP/Keck collaboration (BK15) [205].¹⁶ One may note that a simple quadratic large-field inflation model is ruled out by this dataset as it predicts a too-large value of r , while α -attractors models [207, 208], hilltop inflation [151], or the Starobinsky model [17] (R^2 inflation in Fig. 1.5) can achieve such a small value of r .¹⁷ The model of natural inflation [59] in Fig. 1.5 will be relevant in the context of axion inflation.

In particular, models that fall in the category of small-field inflation fare better at explaining the low value of r , which is implied by the ‘‘Lyth bound’’ [209, 210]: From Eq. (1.31), one can infer the relationship $d\varphi/dN \simeq \sqrt{2\varepsilon_V} M_{\text{P}}$. Thus, for ε_V increasing with φ (decreasing with N), one obtains [151]

$$\frac{\Delta\varphi}{M_{\text{P}}} \gtrsim \left(\frac{r}{10^{-3}} \right)^{1/2} \frac{N}{60}, \quad (1.90)$$

Therefore, models with small field excursions $\Delta\varphi$ generically favor small r .

¹⁶Similar figures for other data sets can be e.g., found in Fig. 5 in Ref. [206] and Fig. 10 in Ref. [33].

¹⁷Note that the combined fit for Planck+ACT+BICEP/Keck+DESI II implies a mildly higher value for n_s , and thus disfavors some of these model, like the Starobinsky model, see Fig. 10 in Ref. [33]. However, there are tensions between the datasets in question, and it is currently not evident that the combined result is reliable. This is however only marginally relevant for this thesis, and we rely on the original PR4 results given at the start of this section.

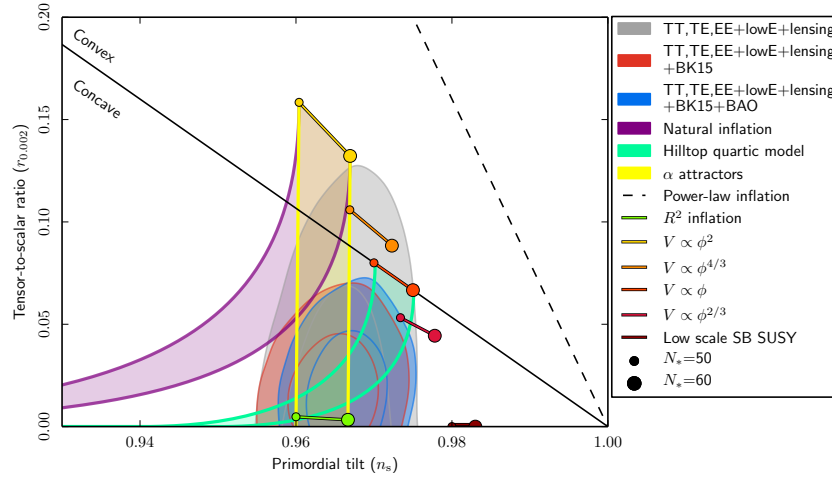


FIGURE 1.5: Constraints on various established slow-roll potentials from PR3 data [14]. Gray (PR3 only), red (PR3+BK15), and blue (PR3+BK15+BAO) contours give 68% and 95% C.L. regions for n_s and r at a pivot scale $k = 0.002 \text{ Mpc}^{-1}$. Theoretical predictions for selected inflationary models are shown as indicated in the right-hand side panel. *This figure is taken from Ref. [14].*

Primordial non-Gaussianity

In Sec. 1.4.1 we have argued that the perturbations from single-field slow-roll inflation are Gaussian. However, this is not entirely true. In fact, even for single-field slow-roll inflation, the spectrum of scalar perturbations mildly deviates from a Gaussian when computing their spectrum to second order in perturbation theory (deviations are of the order of ε_V, η_V) [211, 212]. However, in more elaborate models of inflation, e.g., in models of k-flaton or multifield inflation, the degree of non-Gaussianity in the scalar sector can be more pronounced (see e.g., Refs. [213–217] for reviews). Planck data constrains the degree of primordial non-Gaussianity (PNG) in scalar perturbations, and thereby provides important information on models that go beyond minimal slow-roll models. A comprehensive analysis of PNG in PR3 data is given in Ref. [15]. As we shall see in chapter 3, gauge-field production in axion inflation also induces PNG and is therefore constrained by Planck data. We discuss PNG in axion inflation more in chapter 8.

1.5.3 Signatures beyond the CMB

We have seen how the large physical scales accessible to us via the CMB hold a wealth of information about the physics at early times in the inflationary evolution. However, even in simple single-field slow-roll models, the inflaton potential could deviate significantly from the shape constrained by the CMB, especially when approaching the end of inflation. To probe the physics at such later times during inflation, one needs to detect the signatures that primordial perturbations can leave on scales smaller than those probed by the CMB. To close this chapter, we present two important possibilities which are also relevant for axion inflation, as we discuss in chapter 3.

Gravitational waves from inflation

A promising possibility to probe the physics of inflation at late times is through the direct detection of a stochastic gravitational wave background (SGWB) in current or next-generation GW observatories (see Refs. [218, 219] for reviews). In particular, pulsar-timing array (PTA) and interferometer experiments can collectively cover a vast range of frequencies in GWs: PTAs such as

NANOGrav [220], EPTA [221], or the planned SKA [222] probe nano-Hertz frequencies, ground-based interferometers like LIGO [223, 224], Virgo [225] and KAGRA [226, 227] are sensitive to kilo-Hertz GWs, while intermediate frequencies could in the future be covered by space-born interferometers such as LISA [228], BBO [229] or DECIGO [230]. Collectively, these detectors are therefore sensitive to physical scales that are roughly 9 – 21 orders of magnitude smaller than those accessible via the CMB. Thus, by measuring GWs from inflation, one could probe the physics of inflation up to 50 e -folds after the scales relevant to the CMB have left the horizon.

However, minimal single-field slow-roll inflation predicts an amplitude in IGWs which is likely too low to be detectable by current or next-generation GW experiments. The reason is the low tensor-to-scalar ratio r measured at CMB scales, and the relation $n_t = -r/8$ (see Eq. (1.84b)), which together imply a low amplitude of GWs at CMB scales which only decreases when going to smaller scales, thus higher frequencies.

Alternative scenarios however exist in which the amplitude of IGWs can be enhanced after CMB scales have left the horizon. For example, a blue-tilted GW spectrum can be achieved in non-minimal extensions of slow-roll inflation including, e.g., modifications to gravity [231, 232], or to the kinetic term of the inflation field [233, 234], or even more exotic scenarios [235]. Alternatively, non-standard periods of expansion with equations of states that differ from that of the usual radiation-dominated era can blue-tilt the IGW signal from single-field slow-roll inflation [236–238]. Current measurements by the LIGO–Virgo–KAGRA (LVK) network find no evidence for an SGWB in their frequency range [239, 240], which can be used to constrain such models [241]. Moreover, recent PTA measurements have reported varying evidence in support of an SGWB at nano-Hertz frequencies [242–246]. In particular, the NANOGrav collaboration have studied the possibility of explaining this signal with a blue-tilted spectrum of IGWs [9, 10].

Alternatively, the spectrum of IGWs can be enhanced by a direct source of GWs that is active during inflation. For instance, enhanced small-scale scalar perturbations can source such GWs at second order in perturbation theory, which could result in a detectable signal of *scalar-induced gravitational waves* (SIGWs) [247, 248] (see Ref. [249] for a review on SIGWs). In chapter 9, we study the possibility of an enhanced GW signal sourced by gauge fields during axion inflation.

Primordial black holes

The idea of primordial black holes (PBHs) was first brought forth in Refs [250, 251], the authors of which showed that the gravitational collapse of overdense regions in the early Universe could result in the formation of much lighter or much heavier black holes than those originating from stellar collapse. This motivated the study of PBHs as a candidate for dark matter [91–93] or as the seeds of supermassive black holes in the centers of galaxies [252].

The formation of overdense regions and the collapse to PBHs can be induced when inflationary scalar perturbations reenter the horizon in the radiation-dominated era (see e.g., Ref. [253] for a review). For this formation to yield an appreciable amount of PBHs, one however typically requires a scalar power spectrum that is strongly peaked, with a typical benchmark amplitude being [253]

$$\mathcal{P}_\zeta \sim 10^{-2}. \quad (1.91)$$

However, given that the CMB dictates that $\mathcal{P}_\zeta(k_{\text{CMB}}) \simeq 10^{-9}$ and $n_s < 1$, one therefore does not expect significant PBH formation from standard slow-roll inflation. Seeing as $\mathcal{P}_\zeta \propto |\dot{\phi}|^{-2}$, one way to increase $\mathcal{P}_\zeta$ for $k_{\text{pbh}} > k_{\text{CMB}}$ is by having a period of ultra-slow-roll (USR) inflation on a nearly flat region in the inflaton potential. However, other possibilities exist (see references within Ref. [253]). Moreover, an enhanced scalar power spectrum that can lead to PBH formation is also possible in axion inflation, and we comment further on this possibility in chapter 8.

PBH formation can be used to constrain inflationary models. One way is to require that PBH formation does not overproduce DM (see Refs. [254–256] for reviews). Also, there are two interesting avenues for testing PBHs using GWs (see also Ref. [257] for a review). Firstly, one may directly consider the GW background from PBH mergers, see e.g., the analysis by the LVK collaboration in Ref. [241]. Secondly, PBHs also have an interesting complementarity to SIGWs [258, 259], since both necessitate a large amplitude in scalar perturbations. Combined analyses of SIGWs and PBH constraints have been carried out both by the LVK collaboration [241] and the NANOGrav collaboration [9].

Chapter 2

En route to axion inflation

In the previous chapter, we have established a foundational understanding of slow-roll inflation, which we shall expand upon in the next chapter by considering slow-roll models that identify the inflaton with an axion field. To provide background to this discussion, this chapter gives a first broad introduction to axions from a particle-physics perspective, which then serves to inform our discussion of these fields in the context of inflation in the next chapter.

From a modern perspective, axions constitute a larger class of fundamental fields whose defining properties are that they are nearly massless pseudoscalars that interact weakly with other matter fields (for reviews, see, e.g., Refs. [73–77]). In particular, the lightness of an axion is ensured by an approximate shift symmetry, which is only broken by nonperturbative quantum effects. This same symmetry also restricts the possible interactions of axions with other matter fields to be classically shift-symmetric. From a high-energy-theory perspective, these axions can, for instance, arise as the pseudo-Nambu–Goldstone bosons of a spontaneously broken global symmetry, or they can generically emerge within the context of string theory.

In what follows, we shall unravel the meaning of these different aspects of an axion’s phenomenology by mainly focussing on the original axion field proposed by WEINBERG, WILCZEK, PECCEI and QUINN [69–72]. This axion was introduced to resolve the strong CP problem of quantum chromodynamics (QCD), and may also be referred to as the *QCD axion* to distinguish it from other, more generic, axions, which are often called *axion-like particles* (or ALPs). In particular, it needs to be emphasized that the QCD axion is *not* the inflaton field in axion inflation. Regardless, discussing the QCD axion first proves to be instructive, as many of its properties will also be shared by the axion fields that are relevant in axion inflation models.

For our discussion, we shall first consider how the QCD axion emerges within the context of the strong CP problem in Sec. 2.1. Thereafter, we take a closer look at the QCD axion from a more phenomenological perspective in Sec. 2.2. As part of this presentation, we also establish basic properties of an axion field, e.g., their typical potential and the possible interactions they may have with other matter fields. Finally, we close this chapter by giving an overview of the canonical high-energy physics embeddings of axions in Sec. 2.3.

2.1 A detour to axions

In this first section, we take a rather extensive detour to identify the strong CP problem, and to then introduce the QCD axion as its canonical resolution.

On the surface, the strong CP problem appears as one of several fine-tuning issues in the Standard Model (SM) of particle physics. The QCD Lagrangian allows for a CP-violating term proportional to a phase θ_{QCD} , which would generically induce a non-zero electric dipole moment (EDM) for the neutron. Experimental measurements, however, are consistent with a vanishing neutron EDM, implying that the physical CP-violating parameter of QCD is extremely small. The

strong CP problem is, however, more subtle than this. The parameter θ_{QCD} is deeply connected to the nontrivial vacuum structure of non-Abelian gauge theories such as QCD. This rich topological structure is most clearly exhibited in the Euclidean formulation of Yang–Mills theory, where finite-action field configurations known as instantons interpolate between inequivalent classical vacua. Their existence leads to the identification of the true quantum vacua of Yang–Mills theory, the θ -vacua, which are labelled by a continuous parameter θ . In QCD, this parameter combines with the phase of the quark mass matrix to form the physical angle $\bar{\theta}$, and the strong CP problem is the question of why $\bar{\theta}$ is so close to zero despite there being no clear theoretical reason for it to vanish.

The QCD axion then serves as a possible answer to this mystery. Effectively, this field behaves as a dynamical equivalent to $\bar{\theta}$, and can dynamically relax this angle to zero. The mechanism by which this occurs also gives a mass to the QCD axion. Ultimately, it is this process which proves to be an essential cornerstone of the phenomenology of the QCD axion, and axion fields in general, and which we seek to arrive at in this section. However, to understand this mechanism, we first need to follow the different steps that lead to the strong CP problem in the first place. This means we need to understand the nontrivial vacuum structure of YM theory, identify the θ -vacua, and understand their connection to the angle $\bar{\theta}$. Once this is established, we can then finally introduce the QCD axion, and learn how it attains a mass through interactions with YM instantons, while simultaneously noting how it resolves the strong CP problem.

As our basis for this discussion we rely on two classic textbooks on quantum field theory (QFT) [260, 261], and some additional insights on instantons are taken from Refs. [262, 263]. The discussion of the strong CP problem and the QCD axions also draws on several review articles on axions [73, 76, 77].

Conventions: Let us briefly state the necessary definitions for our subsequent discussion. In our convention, the action for Yang–Mills theory in Minkowski space is,

$$S = -\frac{1}{4} \int d^4x G_{\mu\nu}^a G^{a\mu\nu}. \quad (2.1)$$

where the field-strength tensor $G_{\mu\nu}^a$ is given by

$$G_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + gf^{abc} A_\mu^b A_\nu^c, \quad (2.2)$$

with g the gauge coupling, A_μ^a the gauge field, and f^{abc} the structure constants which we take as real. Furthermore, we define the Lie-algebra-valued gauge potential as $A_\mu = A_\mu^a T^a$, with T^a Hermitian matrices representing the generators of $SU(N)$.

The Euclidean Yang–Mills action is obtained via a Wick rotation, $t \rightarrow -it_E$, resulting in

$$S_E = \frac{1}{4} \int d^4x G_{ij}^a G^{a ij}. \quad (2.3)$$

with the usual identification $iS \equiv -S_E$. Note that we use Latin indices to label the directions in Euclidean four-space, preferably from the middle of the alphabet, $(i, j, k, \dots)$ to avoid confusion with Greek indices which we take from the beginning of the alphabet, $(a, b, c, \dots)$.

2.1.1 The non-trivial vacuum structure of Yang–Mills theory

In a first step towards the strong CP problem, we need to identify the θ -vacua of a YM theory like QCD. The key to doing so is to seek the answers to two related questions:

1. What gauge-field configurations constitute the vacuum configurations of YM?
2. Are the corresponding vacuum states eigenstates of the Hamiltonian H ?

The first question is answered by considering which gauge configurations classically minimize the energy of our theory. In YM theory, the relevant gauge configurations are those for which the field-strength tensor vanishes, $G_{\mu\nu}^a = 0$. The corresponding gauge-field configurations are however not simply described by $A_\mu = 0$. Rather, they are non-trivial pure gauge configurations, $A_\mu^{(n)}$, which are topologically classified in terms of an integer winding number n . Going to the quantum theory, one can then associate to each $A_\mu^{(n)}$ a vacuum state $|n\rangle$.

The second question can be addressed by considering transition amplitudes between two of these vacuum states $|n\rangle$ and $|m\rangle$. One can express such a transition amplitude using the path integral, which sums over all possible gauge configurations that approach the configurations $A_\mu^{(n)}$ at $-T/2$ and $A_\mu^{(m)}$ at $T/2$:

$$\langle m | e^{-HT} | n \rangle \propto \int \mathcal{D}A e^{-S_E[A]}. \quad (2.4)$$

The path integral on the right-hand side is dominated by those classical trajectories that obey its boundary conditions, while also having finite Euclidean action. These solutions are the instantons. If one can identify such instantons in YM theory, one may therefore deduce that there is a non-zero tunneling probability between the states $|n\rangle$ and $|m\rangle$, implying that they cannot be eigenstates of the Hamiltonian. This observation then allows one to identify the true vacua of YM theory as superpositions of $|n\rangle$.

Following this logic, we go to Euclidean space to consider if Yang–Mills theory admits instanton solutions that asymptotically approach vacuum configurations. From Eq. (2.3), one can surmise that any field configuration with finite action needs to fall off sufficiently fast towards the boundary of Euclidean space, i.e.,

$$G_{ij}^a(x) \rightarrow 0 \quad \text{for } |x| \rightarrow \infty, \quad (2.5)$$

The corresponding condition on the gauge field A_i is

$$A_i(x) \rightarrow \frac{i}{g} U^{-1}(\hat{x}) \partial_i U(\hat{x}) \quad \text{for } |x| \rightarrow \infty, \quad (2.6)$$

where $U(\hat{x})$ is an element of $SU(N)$ with $\hat{x} = x/|x|$ to ensure that Eq. (2.6) is preserved. Once again, the asymptotic configuration approached by A_i in Eq. (2.6) has minimal energy, and one can therefore identify it as a pure-gauge vacuum configuration that we previously remarked upon.

The notion of winding number now enters by realizing that the asymptotically pure gauge configurations are defined by a map $U : S^3 \rightarrow SU(N)$. Not every choice of U corresponds to a distinct pure gauge configuration, as one can always perform a continuous gauge transformation to redefine U without actually obtaining a physically different configuration. Put differently, only maps that are not continuously transformable into each other yield inequivalent pure gauge configurations as $|x| \rightarrow \infty$. Mathematically, this means one can classify the asymptotic states in Eq. (2.6) in terms of their homotopy class in $\pi_3(SU(N)) \cong \mathbb{Z}$ [264], so every pure gauge configuration can indeed be classified by an integer n , and we can associate them with a vacuum state $|n\rangle$.

In fact, any finite-action gauge configuration can be classified by an integer, which is encoded in the following gauge-invariant combination

$$G_{ij}^a \tilde{G}^{a ij} = \epsilon^{ijkl} \partial_i \left(A_j^a G_{kl}^a - \frac{g}{3} f^{abc} A_j^a A_k^b A_l^c \right) \equiv \partial_i K^i, \quad (2.7)$$

where $\tilde{G}^{a ij} = \epsilon^{ijkl} G_{kl}^a / 2$ is the Hodge dual to the field-strength tensor. Note that $G_{ij}^a \tilde{G}^{a ij}$ is a total derivative of a vector K^i , which we interpret as a current, whose associated charge can be computed by integrating $G_{ij}^a \tilde{G}^{a ij}$ over $\mathbb{R}^4$. The result is then given by¹

$$\int d^4x G_{ij}^a \tilde{G}^{a ij} = \int d^4x \partial_i K^i = \lim_{|x| \rightarrow \infty} \int_{S_3} d\sigma_i K^i = \frac{32\pi^2 \nu}{g^2} \equiv Q_\nu, \quad (2.8)$$

with ν an integer, which is simply the winding number for a pure gauge configurations.

An important result for the topological charge is the so-called *Bogomoln'yi inequality*, which gives a lower bound on the Euclidean action for any finite-action gauge configuration with topological charge Q_ν , [261, 265],

$$S_E[A] \geq \frac{8\pi^2 |\nu|}{g^2}. \quad (2.9)$$

One implication of this inequality is that configurations with $G_{\mu\nu}^a = 0$, but non-trivial winding n , cannot be globally assumed, as their action is strictly vanishing, in violation of Eq. (2.9). However, solutions to the classical Euclidean equations of motion can be found which saturate the bound in Eq. (2.9), namely the instantons of YM theory. Explicit instantons with $\nu = \pm 1$ are derived in the seminal works by BELAVIN ET AL. [265]. These field configurations are called BPST instantons, and their associated action is $S_E = 8\pi^2/g^2$.

Let us then consider the physical interpretation of an instanton configuration A_i with topological charge Q_ν and finite action $S_E = 8\pi^2 |\nu|/g^2$. According to Eq. (2.6), A_i necessarily needs to approach a pure gauge configuration as $|x| \rightarrow \infty$. To simplify our argument, we use our gauge freedom to set $A_0 = 0$ such that $A_i(|x| \rightarrow \infty)$ is independent of the time-like direction $\hat{x}_0$ (see Eq. (2.6)). We therefore have that A_i approaches a time-independent pure gauge for $T \rightarrow \pm\infty$, each with definite winding number $n_\pm$. We can then use Eq. (2.8) to find a relationship between the winding numbers of the asymptotic states and the topological charge of A_i , namely

$$\nu = \frac{g^2}{32\pi^2} \int d^4x G_{ij}^a \tilde{G}^{a ij} = \frac{g^2}{32\pi^2} \int d^4x \partial_0 K^0 = \frac{g^2}{32\pi^2} \int d^3\mathbf{x} K^0 \Big|_{T=-\infty}^{T=\infty} = n_+ - n_-,$$

where we used that $A_0 = 0$ means $K_i = 0$ for $i \in \{1, 2, 3\}$. Therefore, the integer ν labelling the topological charge of an instanton gives the change in winding number effected by it when going from one asymptotic pure gauge configuration to another.

In light of this observation, the BPST instantons with topological charge $Q_{\pm 1}$ can therefore be understood as changing the winding number between two separate pure gauge configurations at $T \rightarrow \pm\infty$ by ± 1 . In light of our discussion following Eq. (2.4), we can therefore conclude that the existence of BPST instantons implies a non-zero transition probability between vacuum states $|n\rangle$ and $|n \pm 1\rangle$. We are thus left to conclude that the $|n\rangle$ states cannot be the true vacuum state of YM theory. Instead, the true vacuum states are rather the so-called θ -vacua, which were first

¹For an elegant derivation of why the integration over this current yields the winding number for a pure gauge configuration, see Ref. [261].

identified by CALLAN, DASHEN AND GROSS [266], and JACKIW AND REBBI [267],

$$|\theta\rangle = \sum_n e^{in\theta} |n\rangle. \quad (2.10)$$

Here θ is an angular parameter which can take any value in $(-\pi, \pi)$.

2.1.2 Much ado about θ

Above, we sketched the basic line of reasoning which leads to the identification of the θ -vacua. What we should now consider are its implications, as these bring us closer to the strong CP problem.

We start by considering the expectation value of an arbitrary operator $\mathcal{O}$ with respect to $|\theta\rangle$. Using the notion of topological charge, we can express this in terms of the Euclidean path integral as

$$\langle\theta|\mathcal{O}|\theta\rangle = \sum_{n,m} e^{i(n-m)\theta} \langle m|\mathcal{O}|n\rangle = \frac{\sum_\nu \int [\mathcal{D}A]_\nu \mathcal{O}[A] e^{-S_E[A] + i\nu\theta}}{\sum_\nu \int [\mathcal{D}A]_\nu e^{-S_E[A] + i\nu\theta}}, \quad (2.11)$$

where the integral measure $[\mathcal{D}A]_\nu$ is to be understood as an integration over individual gauge sectors with fixed topological charge Q_ν with $\nu = n - m$. Note the effect of the θ parameter, which adds a weight to the individual topological sectors, $e^{i\nu\theta}$. To simplify the treatment of this term, one can use the definition of the topological charge in Eq. (2.8) to express ν in terms of $G_{ij}^a \tilde{G}^{a ij}$. The exponent in Eq. (2.11) can then be absorbed into a new Euclidean action describing our theory,

$$S_{E,\theta}[A] \equiv S_E[A] - i\nu\theta = S_E[A] - i\theta \frac{g^2}{32\pi^2} \int d^4x G_{ij}^a \tilde{G}^{a ij}. \quad (2.12)$$

Note that the combination $G\tilde{G}$ is odd under a CP transformation. Therefore, the theory described by $S_{E,\theta}[A]$ explicitly violates CP for $\theta \neq 0$, despite the fact that the original action, $S_E[A]$, did not. We can thus see first hints of CP violation appearing that are associated with θ .

The representation of the path integral as in Eq. (2.11) can also be used to compute the ground-state energy, $E(\theta)$, for a given θ -vacuum. This result will play a significant role when we later discuss the phenomenology of axions and the resolution of the strong CP problem. A simple result for $E(\theta)$ can be computed via a semi-classical computation technique known as the *dilute instanton-gas approximation* (DIGA), which was first applied in the seminal works on instantons by 't HOOFT [268], and later discussed in the context of QCD in Ref. [269].

The strategy to compute $E(\theta)$ is to consider the transition amplitude

$$\langle\theta|e^{-HT}|\theta\rangle \propto \sum_\nu \int [\mathcal{D}A]_\nu e^{-S_E[A] + i\nu\theta}. \quad (2.13)$$

in the limit $T \rightarrow \infty$, and to evaluate the right-hand side in a semi-classical approach. In this limit, the result one obtains is dominated by the ground-state energy, as any other terms are exponentially suppressed. The DIGA relates to the strategy in which the semi-classical calculation is performed. Each sector labelled by ν takes its dominant contribution from an instanton with charge Q_ν . Since we know that instantons with charge $Q_{\pm 1}$ exist, and because topological charge is additive, the idea is then to describe the instantons with charge Q_ν as a collection of n instantons with charge Q_1 and $\bar{n}$ anti-instantons with charge Q_{-1} , such that $\nu = n - \bar{n}$. As each individual (anti-)instanton contributes $S_0 = 8\pi^2/g^2$ to the action, an instanton with charge Q_ν therefore contributes in total $S_\nu = S_0(n + \bar{n})$ to the above path integral, and the sum over ν can be expressed as a sum over individual BPST instantons of charge $Q_{\pm 1}$. The approximation further relies on the fact

that individual instantons are well-localized with a finite spatial extent, something which can be explicitly seen from the explicit BPST solutions, for example given in Refs. [260–262].

The result of this procedure is

$$E(\theta) \propto -\exp(-8\pi^2/g^2) \cos \theta. \quad (2.14)$$

This explicitly demonstrates the fact that all θ -vacua are distinct from one another as they all are associated with a different ground-state energy. It is important for our later discussion of axions, so we should discuss it a bit further.

The first thing to note about Eq. (2.14) is that the energy $E(\theta)$ is a nonperturbative effect, as it vanishes in the limit of zero gauge coupling. The reason for this is that the instantons themselves are nonperturbative phenomena, as can be seen by their action depending on inverse powers of g . Therefore, the effects of the θ -vacuum are only important at energy scales where $g \sim \mathcal{O}(1)$ (e.g., for QCD below $\Lambda_{\text{QCD}} \sim 150 \text{ MeV}$).² Other notable features of $E(\theta)$ are its periodicity, symmetry under the exchange $\theta \rightarrow -\theta$, and that it is minimal for $\theta = 0$. These properties are in fact generic to $E(\theta)$, and do not depend on the DIGA computation [77]: The periodicity of $E(\theta)$ is a consequence of ν in Eq. (2.11) being an integer, and the Vafa–Witten theorem [270] states that parity cannot be spontaneously broken (for a vector theory like QCD [271]), implying $E(0) \leq E(\theta)$. In addition, if $S_{E,\theta=0}$ is CP invariant, $E(\theta) = E(-\theta)$ since only the term $G\tilde{G}$ in Eq. (2.12) violates CP, and a CP transformation then only amounts to changing the sign of θ .³ However, it should be noted, that, despite its successes, the result in Eq. (2.14) has a major flaw, which is however hidden behind the proportionality sign; the pre-factor we did not write down contains a divergent integral over the size of the single instantons.⁴ In particular, this means that Eq. (2.14) can merely be used as a qualitative result, which is however sufficient for our purposes. Alternatively, in QFT at finite temperatures, the DIGA approach becomes valid again, as the thermal mass of the non-Abelian gauge boson regularizes the divergent pre-factor [273].

With this result for $E(\theta)$, we have established all that we need to know about the θ vacuum for pure Yang–Mills theory to understand the strong CP problem, and ultimately the role of the QCD axion. As a last part of this analysis, we should then move from Euclidean space back to Minkowski space via a Wick rotation. Performing these steps starting from Eq. (2.12), and identifying the Minkowski action via $iS_\theta \equiv -S_{E,\theta}$, one gets

$$S_\theta[A] = S[A] + \theta \frac{g^2}{32\pi^2} \int d^4x G_{\mu\nu}^a \tilde{G}^{a\mu\nu}. \quad (2.15)$$

where, just as in the Euclidean theory, the θ parameter appears as regulating the strength of the CP violating term $G\tilde{G}$.

θ in the presence of fermions

We again see first glimpses of the CP violation appearing which ultimately will lead to the strong CP problem. However, so far we have just considered pure Yang–Mills theory, but the strong CP problem we are seeking to understand is relevant for QCD, which also contains quarks. We should then consider how the CP violation through θ interacts with fermions in the theory. Therefore, we

²The full result for $E(\theta)$ requires renormalization, and depends on the running coupling $g(\mu)$ [260].

³Note that, these considerations do not hold exactly in the Standard Model (SM) [77]. Because the SM is a chiral theory, its true minimum of E is shifted by a small amount. Furthermore, explicit CP violation in the CKM phase means that there is a small odd contribution to $E(\theta)$.

⁴An interested reader can find the explicit pre-factor in Refs. [260, 262], or the original articles in which it was computed, e.g., Ref. [268] for the case of an $SU(2)$ or Ref. [272] for $SU(N)$.

now add a family of Dirac fermions $\psi = (\psi_1, \psi_2, \dots)$ charged under the gauge group associated with $G_{\mu\nu}^a$ to $S_\theta[A]$ in Eq. (2.15). The connecting piece between the θ -term and the fermions ψ are chiral rotations, which we shall now briefly review following Refs. [261, 274]

First, some general things. If we perform a chiral rotation of ψ

$$\psi(x^\mu) \rightarrow \exp(i\gamma^5 \alpha(x^\mu) \tau) \psi(x^\mu), \quad (2.16)$$

with τ some Hermitian matrix acting on ψ , this affects the path-integral of our theory by changing the measure of ψ [275]

$$[\mathcal{D}\psi][\mathcal{D}\bar{\psi}] \rightarrow \exp\left(i \frac{g^2}{8\pi^2} \int d^4x \left[\alpha G_{\mu\nu}^a \tilde{G}^{b\mu\nu} \right] \text{Tr}(T^a T^b \tau)\right) [\mathcal{D}\psi][\mathcal{D}\bar{\psi}], \quad (2.17)$$

where T_a denotes the generators of $SU(N)$ in the representation associated with ψ . The additional terms in the exponent effectively constitute new terms in our action describing fermions and gauge-fields. In fact, if the chiral rotation in Eq. (2.16) is a symmetry of this action, this change in the path integral measure is the same as the chiral anomaly of the associated classical conserved current J_5^μ [276, 277];

$$\partial_\mu J_5^\mu = -\frac{g^2}{8\pi^2} \alpha G_{\mu\nu}^a \tilde{G}^{b\mu\nu} \text{Tr}(T^a T^b \tau). \quad (2.18)$$

With these generalities established, we now restrict ourselves to a global chiral transformation, $\psi_f \rightarrow \exp(i\gamma^5 \alpha_f) \psi_f$, where f labels the individual fermion flavors in ψ . Eq. (2.17) implies that this rotation leads to the appearance of a new term proportional to $G\tilde{G}$ in the action, which amounts to effectively redefining the θ angle in Eq. (2.15) to

$$\theta \rightarrow \theta + 2 \sum_f \alpha_f. \quad (2.19)$$

We are thus tempted to think that we can remove θ via an arbitrary chiral rotation, so that it does not have a physical effect. However, this depends on if the chiral rotation $\psi_f \rightarrow \exp(i\gamma^5 \alpha_f) \psi_f$ leaves the action invariant, i.e., if it is a symmetry of the action. If so, the chiral rotation can be simply used to remove θ , the resulting theory still describes the same physics, and θ has no meaningful effect. However, things are different if the chiral transformation is not a symmetry of the action.

Consider for example the case where the fermions ψ_f have a mass m_f , which breaks the chiral symmetry. A chiral rotation $\psi_f \rightarrow \exp(i\gamma^5 \alpha_f) \psi_f$ amounts to a change

$$m_f \rightarrow m_f \exp(2i\alpha_f). \quad (2.20)$$

At the same time, the chiral transformation rotates the θ -angle as in Eq. (2.19). Since the only thing the rotation achieves is a change of variables in the path integral, the underlying physics should not be affected by it. Thus, any physical observable can only depend on the combination

$$\exp(-i\theta) \prod_f m_f. \quad (2.21)$$

In particular, if the chiral symmetry in question is explicitly broken by massive fermions, one can always perform a global chiral rotation to set $\theta = 0$, but this would simply re-introduce explicit CP violation in the phases of the fermion masses. More generally, the additional CP violation associated with a θ -term cannot be removed by a chiral rotation which is not a symmetry of the classical theory.

2.1.3 The strong CP problem

We now have all the ingredients we require to describe the strong CP problem. QCD is a non-Abelian gauge theory and thus permits a θ angle, namely θ_{QCD} .⁵ To see if θ_{QCD} is physically relevant, we need to assess if QCD has any classical chiral symmetries by which it can be set to zero. In fact, QCD has a chiral symmetry for each massless quark. But unfortunately, none of the quarks appear to be massless; lattice computations indicate that the lightest quark, the up-quark, has a mass of about 2 MeV [278]. Therefore, there is no chiral symmetry by which θ_{QCD} can be removed. According to Eq. (2.21), physical observables will then depend on a combination of mass parameters and θ_{QCD} , which, for the Standard Model (SM), is

$$\bar{\theta} = \theta_{\text{QCD}} + \text{Arg det } Y_U Y_D, \quad (2.22)$$

where Y_U and Y_D are the Yukawa matrices for up- and down-type quarks.

The most prominent observable affected by $\bar{\theta}$ is the electric dipole moment of the neutron (nEDM), d_n , which describes the interaction strength between an electric field and the neutron's spin. Experimental measurements however indicate that d_n is consistent with zero,

$$|d_n^{\text{exp}}| < 1.8 \times 10^{-26} \text{ e cm}. \quad (2.23)$$

When compared to the theoretical predictions for d_n [279–287], this leads to the conclusion

$$|\bar{\theta}| \lesssim 10^{-10}. \quad (2.24)$$

This unexplained smallness of $\bar{\theta}$ constitutes the strong CP problem.

The canonical resolution of this puzzle is the introduction of a pseudoscalar axion field, which we discuss in short. However, alternative proposals have also been put forth. Several authors have suggested that the UV completion of the SM preserves parity or CP to address the strong CP problem [288–292], but these models typically requires some fine-tuning to work (see, e.g., Refs [293, 294]). Instead of seeking the answer in the UV completion, other authors instead propose unknown IR effects in QCD. For example, Refs. [295–298] propose additional spacetime dimensions that trivialize the QCD vacuum, while Refs. [299, 300] invoke screening effects due to confinement. Finally, Ref. [301] claimed that the strong CP problem is in fact a mathematical artifact of its derivation, and that no physical observable, including the nEDM, can depend on the CP violation from θ -terms. Recently, these claims have been refuted [302], but the discussion is ongoing [303].

2.1.4 Introducing the QCD axion

Let us then, finally, introduce the QCD axion, and see how it addresses the strong CP problem. To show this, we take for simplicity a light-weight version of QCD described by the Lagrangian,

$$\mathcal{L}_{\psi G} = \bar{\psi}(i\gamma^\mu D_\mu - m_\psi)\psi - \frac{1}{4}G_{\mu\nu}^a G^{a\mu\nu} + \theta \frac{g^2}{32\pi^2} G_{\mu\nu}^a \tilde{G}^{a\mu\nu}, \quad (2.25)$$

where we have just included a single quark ψ as a Dirac fermion, with m_ψ its mass. Following our previous discussion, we also include a θ -term, which we define to contain all sources of CP violation, including any possible CP violation from m_ψ .

⁵Should we include a θ term for the $SU(2)_L$ gauge group of the SM? Yes, but it can be removed via a chiral rotation which preserves total baryon and lepton number.

The QCD axion a is now introduced as a pseudoscalar field which couples to gluons in the same way as the θ term does.⁶ Its introduction amounts to adding the following terms to $\mathcal{L}_{\psi G}$,

$$\mathcal{L}_{a\psi G} = \mathcal{L}_{\psi G} + \frac{1}{2} \partial_\mu a \partial^\mu a - \frac{a}{f} \frac{g^2}{32\pi^2} G_{\mu\nu}^a \tilde{G}^{a\mu\nu} \quad (2.26)$$

where the energy-scale f , the so-called *axion decay constant*, is introduced to ensure the correct mass dimensions of the axion–gauge-field interaction. For the moment, we can see it as indicating some high-energy completion whose effects are suppressed below f . Note that a is necessarily a pseudoscalar, since it couples to the CP-odd combination $G\tilde{G}$, and we do not wish to add any new explicitly CP-violating terms to $\mathcal{L}_{\psi G}$. Additionally, we know that $G_{\mu\nu}^a \tilde{G}^{a\mu\nu}$ is classically a total derivative, so a inherently comes with a shift symmetry, $a \rightarrow a + C$. To preserve this symmetry, we do not include a tree-level potential for a .

The key to resolving the strong CP problem lies in the effective potential of the axion. We know that $G\tilde{G}$ is related to the topological charge of the gluon field, so the axion effective potential arises from its interactions with the associated instantons. One approach to determine $V_{\text{eff}}(a)$, is to follow the original work in Refs. [71, 72]. This strategy involves computing $V_{\text{eff}}(a)$ from the effective Euclidean action for $\mathcal{L}_{a\psi G}$ by evaluating it in the presence of a constant background axion field, a_0 . The basic idea is to realize that a_0/f effectively acts in the same way as the θ angle, such that the calculation for V_{eff} can be analogously performed to the one of $E(\theta)$ by using the DIGA. The result is then

$$V_{\text{eff}}(a) = -\frac{1}{VT} \ln \left(\int \mathcal{D}A \mathcal{D}\psi \mathcal{D}\bar{\psi} e^{-S_E[a_0]} \right) \propto -\cos \left(\theta - \frac{a}{f} \right), \quad (2.27)$$

where $S_E[a_0]$ denotes the Euclidean action for $\mathcal{L}_{a\psi G}$, which, for constant a_0 and in the absence of external currents, is the same as the effective Euclidean action. From Eq. (2.27) we can see the mechanism by which the axion resolves the strong CP problem; the axion field minimizes its energy by assuming $a = \theta f$, thereby erasing any CP violation coming from a non-zero θ . Here, it is also worth noting that all considerations that apply to $E(\theta)$ in Eq. (2.14) apply also to Eq. (2.27). In particular, this also means that the DIGA result for $V_{\text{eff}}(a)$ contains a divergence, meaning Eq. (2.27) does not reflect the true axion potential. However, also all other properties translate to $V_{\text{eff}}(a)$, so it is generically periodic, symmetric under an exchange of signs, and minimal for $a = \theta f$. One can therefore always redefine a to absorb the θ angle, $a \rightarrow a - \theta f$, such that the vacuum state of the QCD axion is $a_{\text{min}} = 0$.⁷ These properties ensure that the axion generically addresses the strong CP problem, despite the fact that the effective potential in Eq. (2.27) was derived assuming the DIGA. Additionally, although $V_{\text{eff}}(a)$ breaks the axions shift symmetry, its periodicity means it is merely reduced to a discrete symmetry, $a \rightarrow a + 2\pi f$.

If we then finally take a step back from the strong CP problem, and consider the properties of the QCD axion more abstractly, we can recognize several of the defining features of an axion that we have remarked upon at the start of this chapter: the QCD axion was initially introduced in Eq. (2.26) as a massless pseudoscalar field with a shift symmetry. However, through nonperturbative interactions with the instantons of QCD, it attains a non-zero mass. In this context, an important thing to note is that $V_{\text{eff}}(a)$ can only arise because QCD is confining in the low-energy limit, since, just as $E(\theta)$, the effective potential vanishes for $g \rightarrow 0$. As QCD is confining at a scalar $\Lambda_{\text{QCD}} \sim 150 \text{ MeV}$, this then qualitatively sets the height of the axions effective potential, i.e., $V_{\text{eff}}(a) \sim \Lambda_{\text{QCD}}^4 \cos(a/f)$. In particular, this means its mass is $m_a \simeq \Lambda_{\text{QCD}}^2/f$, making the QCD axion generically light if f is large, precisely as remarked upon at the start of this chapter.

⁶Note that “ a ” does not refer to the scale factor in this chapter.

⁷Keeping in mind the caveats for the SM we mentioned in footnote 3.

2.2 The phenomenology of axions

Now that we have seen how the QCD axion emerges within the context of the strong CP problem, and having identified some of its defining properties, we next take a closer look at the QCD axion to refine some of these ideas. In particular, we also want to learn how the QCD axion interacts with other matter fields. Having done this, we then take a step back to extrapolate the properties of generic axions which are *not* the QCD axion. These properties will then reappear in the next chapter in the context of axion inflation models.

2.2.1 A closer look at the QCD axion

Let us begin with a closer look at the properties of the QCD axion.⁸ Our starting point is a simple low-energy Lagrangian for the QCD axion in the spirit of our “QCD-light” model in the previous section. Taking only two quark flavors into account, the relevant terms concerning the QCD axion are [77]

$$\mathcal{L}_{\text{aQCD}} = \frac{1}{2} \partial_\mu a \partial^\mu a - \frac{a}{f} \frac{g^2}{32\pi^2} G_{\mu\nu}^a \tilde{G}^{a\mu\nu} - (\bar{q}_L M_q q_R + h.c.) \quad (2.28)$$

where, as before, $G_{\mu\nu}^a$ describe the gluons. The quarks are described by the field q , and we have also included a quark mass matrix, $M_q = \text{diag}(m_u, m_d)$.

Starting from a Lagrangian of the type in Eq. (2.28), in the previous section we proceeded to study the effect of the topological term $G\tilde{G}$, which lead us to the effective potential of the QCD axion. However, one can take an illuminating alternative path instead to arrive at interactions between the QCD axion and other matter fields: according to Eq. (2.17), we can remove the gluonic term by performing a chiral rotation of the quarks, $q \rightarrow \exp(i\gamma^5 a/(2f)Q_a)q$ with Q_a satisfying $\text{Tr} Q_a = 1$. In doing so, we need to be careful to also remember the electric charge carried by quarks. The change to the path-integral measure in Eq. (2.17) applies to *all* gauge groups under which the quarks are charged. This means, that a chiral rotation of the quarks generates a new coupling between a and the photon field $F_{\mu\nu}$. Moreover, the quark mass term is not invariant under a chiral transformation, so we also generate a new interaction between the axions and the quarks. In total, the Lagrangian after the rotation reads

$$\mathcal{L}'_{\text{aQCD}} = \frac{1}{2} \partial_\mu a \partial^\mu a + \frac{\alpha N_c}{4\pi f} \text{Tr}(Q_a Q^2) a F_{\mu\nu} \tilde{F}^{\mu\nu} - \frac{\partial_\mu a}{2f} Q_a \bar{q} \gamma^\mu \gamma^5 q - (\bar{q}_L M_a q_R + h.c.) , \quad (2.29)$$

where $M_a = e^{i\frac{a}{2f}Q_a} M_q e^{i\frac{a}{2f}Q_a}$, $\alpha = e^2/(4\pi)$ is the fine-structure constant, $N_c = 3$ is the number of colors, and $Q = \text{diag}(2/3, -1/3)$ is a matrix containing the charges of the up- and down quark.

What we can deduce from Eqs. (2.29) is that the QCD axion generically couples to photons and quarks, even though we did not include these interactions in Eq. (2.28). This is an unavoidable consequence of the fact that the axion needs to couple to gluons to resolve the strong CP problem, and that quarks couple to both gluons and photons. In fact, in an effective-field-theory (EFT) approach, we could have included any such shift-symmetric coupling between the axion, photons, quarks (and also electrons if we like) in Eq. (2.29). However, even if we do not, axion–photon and axion–quark couplings are always present (electron couplings are not, as they are color singlets). Note also that all couplings of the QCD axions to these matter fields are suppressed by $1/f$, and are therefore generically weak. In particular, a large value of f , and thus a weak coupling to matter are enforced from beam-dump experiments which rule out strong axion–photon interactions [304].

Indeed, a large value of f also entails that the QCD axion is light, as we have already noted upon in the previous section. Let us follow up on this idea in a bit more detail. In the low-energy

⁸Our discussion is mainly oriented along the one presented in Ref. [77].

limit of QCD, other techniques than the DIGA can be used to compute $V_{\text{eff}}(a)$, in particular chiral perturbation theory (χ PT), wherein one exchanges gluons and quarks for an effective description in terms of pions. On the ground state of the neutral pion, the axion potential at leading order in χ PT takes the form [77, 305, 306]

$$V_{\text{eff}}(a) = -m_\pi^2 f_\pi^2 \sqrt{1 - \frac{4m_u m_d}{(m_u + m_d)^2} \sin^2\left(\frac{a}{2f}\right)}, \quad (2.30)$$

where m_π is the pion mass and $f_\pi \simeq 92$ MeV is the pion decay constant.

There are a few things we should discuss about Eq. (2.30). Firstly, it explicitly demonstrates the properties of $V_{\text{eff}}(a)$ that we have underpinned in the discussion following Eq. (2.27); as expected, the potential in Eq. (2.30) is not reproduced by the DIGA result. However, the generic features are there: it is a periodic under a shift $a \rightarrow a + 2\pi f$, and its minimum is at $a = 0$. Secondly, note that the energy scales, m_π and f_π , are in good qualitative agreement with our previous estimate of the potential height being Λ_{QCD}^4 . From Eq. (2.30), one finds that the mass of the QCD axion is

$$m_a^2 = \frac{m_u m_d}{(m_u + m_d)^2} \frac{m_\pi^2 f_\pi^2}{f^2} \implies m_a \simeq 5.7 \left(\frac{10^{12} \text{ GeV}}{f} \right) \text{ peV}. \quad (2.31)$$

Indeed, we see that the QCD axion is generically light for typical values $f \sim 10^{12}$ GeV.

The lightness of the QCD axion also means that it is unsuitable as an inflaton candidate, for which one may expect substantially larger masses, in particular for the typical energy scale of inflation, $H_{\text{infl.}} \sim 10^{13}$ GeV (cf. Eq. (1.86)). For a suitable inflaton candidate, a more generic kind of axion field is needed, and we shall briefly deliberate on the properties of such generic axions in the next section. Before we do so, some final remarks on the QCD axion.

While the QCD axion is not a suitable inflaton candidate, it is a prospective dark-matter candidate, as it is generically weakly interacting [66–68]. In particular, the QCD axion is distinctively different from other candidate dark-matter particles such as the conventional weakly-interacting massive particles (WIMPs), for which typical masses are in the TeV mass range [278]. It should also be noted that, although the QCD axion couples weakly to matter, these interactions are not phenomenologically irrelevant. For instance, the axion–photon interaction means that axions can be converted to photons in the presence of strong magnetic fields. This is used in several experimental designs aimed at detecting the QCD axion, e.g., in haloscopes and helioscopes [307], or light-shining-through-wall experiments [308]. Meanwhile, even weak axion–matter interactions, in particular those with electrons or photons, also entail that the axion can affect cosmological or astrophysical environments. For example, stellar evolution would be altered by too efficient axion-cooling processes, while the decay of the QCD axion into photons in the early Universe would have distorted the CMB black-body spectrum [309]. Reviews on cosmological implications of axions are given by Refs. [76, 307], while Refs. [74, 310] focus on axions in astrophysical contexts.

2.2.2 Recipe for a generic axion

Now that we have familiarized ourselves with the QCD axion, we can take a step back to consider what a generic axion field looks like. We have seen in the previous section that an axion is a pseudoscalar, it may obtain a mass via interactions with a confining gauge-field sector, and it couples weakly to other matter fields. To summarize these properties, we can, for instance, use these ingredients to describe generic axions in the language of an EFT, which we consider to be valid below an energy scale f .

To start out, we define the axion as a pseudoscalar ϕ with a tree-level shift symmetry, implying it has no tree-level potential (we can add a constant potential, V_0). Then, we should consider the operators which are allowed by the symmetries of our theory. Restricting ourselves to consider only CP-preserving operators, the lowest-dimensional couplings between ϕ and an arbitrary gauge field $F_{\mu\nu}^a$ is

$$\phi \frac{\alpha_F}{f} F_{\mu\nu}^a \tilde{F}^{a\mu\nu}, \quad (2.32)$$

where α_F is some dimensionless coupling constant. Couplings to fermions are instead of the type

$$\frac{\partial_\mu \phi}{f} c_\psi^0 \bar{\psi} \gamma^\mu \gamma^5 \psi, \quad (2.33)$$

where ψ denotes an arbitrary fermion field and (for simplicity, a Dirac fermion), and c_ψ^0 is again a dimensionless coupling constant.

To generate a mass for the axion without explicitly violating its shift symmetry, a nonperturbative interaction with a confining non-Abelian gauge-field sector is a canonical possibility, as we have seen in Sec. 2.1. Assuming, e.g., that the gauge-field sector becomes confining at some energy scale Λ , a generic axion potential would then be

$$V(\phi) = \Lambda^4 \left[1 - \cos \left(\frac{\phi}{f} \right) \right]. \quad (2.34)$$

Although our discussion of the QCD axion has shown that this likely only approximates a realistic axion potential, it captures all its relevant features, in particular its periodicity. The mass of a generic axion is then

$$m_\phi^2 = \left. \frac{\partial^2 V}{\partial \phi^2} \right|_{\phi=0} = \frac{\Lambda^4}{f^2}. \quad (2.35)$$

In particular, for a generic axion, the energy scale Λ does not need to be associated with the confinement scale of QCD, so a generic axion field can be much lighter, or much heavier than the QCD axion is, depending on the ratio of Λ and f .

We can also consider the coupling of ϕ to fermions and other gauge fields. We are always free to explicitly include interaction terms like Eq. (2.32) or Eq. (2.33) in our effective Lagrangian; these are, after all, permitted by our symmetries. However, we need to remember our discussion of the QCD axion; because the quarks carry color charge, the QCD axion couples to them. Furthermore, because the quarks also couple to photons, so does the QCD axion. The same is then true for our generic axion. For example, consider some fermion field ψ which couples to $G_{\mu\nu}^a$, and another independent Abelian field $F_{\mu\nu}$. Even if the original Lagrangian did not include a coupling between ϕ and ψ initially, we can always make an explicit interaction as in Eq. (2.33) appear via a chiral rotation. By the same rotation, we then also generate an interaction between $F_{\mu\nu}$ and ϕ as in Eq. (2.32), even if it was explicitly present before. In this sense, an arbitrary axion generically couples to fermions, non-Abelian gauge bosons and Abelian gauge bosons.

As we have already alluded to multiple times, axions that are substantially heavier than the QCD axion can be relevant in the context of axion inflation models. Meanwhile, axions that are substantially *lighter*⁹ than the QCD axion have however also been considered for a variety of phenomenologically interesting reasons. Ultra-light axions ($m_\phi \sim 10^{-26} - 10^{-20}$ eV) are often considered as a form of ultralight dark matter (ULDM), whose primary motivation is that it would

⁹Note that “lighter than the QCD axion” is not a clear-cut division, as the mass of the QCD axion depends on f as in Eq. (2.31). Ref. [76] roughly places the dividing line at a mass of 10^{-12} eV, corresponding to $f \sim M_{\text{P}}$. For example, superradiance discussed below could also be relevant to a very light QCD axion with $m_a \sim 10^{-12}$ eV.

address the so-called *small-scale crises* of CDM [311, 312]. ULDM can be probed experimentally either through their imprint in the CMB or large-scale structure [313, 314], or using pulsar-timing arrays [315], as for example recently analyzed in Refs. [316–318]. Even lighter axions ($m_\phi \sim 10^{-30}$ eV) are also considered in the context of dark energy [319], where they fall in the class of *quintessence* models (for a review, see Ref [320]). Another interesting implication of light axions is their interaction with black holes (BHs). By extracting energy and angular momentum from a spinning Kerr BH [321], a bosonic field can form stable orbits around a BH, similar to the electron field in an atom. This effect is known as *superradiance* (for a review, see, e.g., Ref. [322]). A last notable effect is *cosmic birefringence*: if a light axion field coupled to photons permeates our Universe it could cause rotations in the polarization plane of CMB photons [323]. Several recent analyses of different CMB datasets in fact find varying evidence for this effect [324–326].

2.3 The origin of axions

As the last part of our discussion of axions, we now briefly summarize the two most prominent examples for how axion fields arise from the perspective of high-energy physics.

2.3.1 The Peccei–Quinn mechanism

The Peccei–Quinn mechanism (PQ mechanism) [71, 72] proposes that an axion may arise as the Nambu–Goldstone (NG) boson of a spontaneously broken global symmetry, for instance a global $U(1)_{\text{PQ}}$. This entails that the axion field is initially massless at the energy scale where the global symmetry is broken, and only at a lower energy scale it attains a mass via nonperturbative effects, e.g., the interactions with non-Abelian gauge sector, as discussed in previous sections.

The basic premise of the PQ mechanism is best explained by considering an explicit toy model.¹⁰ We start with the Lagrangian

$$\mathcal{L} = \partial_\mu \Phi^\dagger \partial^\mu \Phi + i\bar{\psi}\gamma^\mu D_\mu \psi - V(|\Phi|) - (y_\psi \Phi \bar{\psi}_L \psi_R + h.c.), \quad (2.36)$$

where $\psi = (\psi_L + \psi_R)$ is a Dirac fermion with $\psi_{L/R} = (1 \mp \gamma_5)\psi/2$ its left- and right-handed components. For the PQ mechanism to work, we need to assume that ψ transforms under the fundamental representation of some $SU(N)$ group with N_c colors. For the sake of illustration, we also associate a charge Q of some $U(1)$ gauge group to ψ , although this is not a requirement for our toy model to work. For simplicity, we take the complex scalar Φ to be a singlet with respect to both gauge groups.

To identify the PQ symmetry, note that the Lagrangian in Eq. (2.36) is invariant under the following axial $U(1)_{\text{PQ}}$ transformation,

$$\Phi \rightarrow e^{i\alpha} \Phi, \quad \psi \rightarrow e^{-i\gamma_5 \frac{\alpha}{2}} \psi, \quad (2.37)$$

Classically, Noether’s theorem implies that this symmetry corresponds to a global conserved current, J_{PQ}^μ whose fermionic contributions is given by $J_{\text{PQ},\psi}^\mu = \bar{\psi}\gamma^\mu\gamma_5\psi$, and we say that $\psi_{L/R}$ carry PQ charge $Q_{L/R} = \pm 1/2$.

Now, we consider that the scalar potential is of the typical sombrero-type for spontaneous symmetry breaking,

$$V(|\Phi|) = \lambda_\Phi \left(|\Phi|^2 - \frac{v_\phi}{2} \right)^2. \quad (2.38)$$

¹⁰For this discussion, we synthesize ideas presented in Ref. [77] and also take some insights from Ref. [327].

The field Φ thus assumes its vacuum expectation value v_ϕ , thereby spontaneously breaking $U(1)_{\text{PQ}}$. Writing Φ as

$$\Phi = \frac{1}{\sqrt{2}}(v_\phi + \rho)e^{i\phi/v_\phi}, \quad (2.39)$$

the radial field ρ attains a mass, $m_\rho = \sqrt{2\lambda_\Phi}v_\phi$, and ϕ is the massless NG boson of the spontaneously broken PQ symmetry. In particular, ϕ enjoys a shift symmetry, and we may later identify it as an axion.

Through its Yukawa interaction with Φ , ψ also attains a mass, $m_\psi = y_\psi v_\phi / \sqrt{2}$;

$$(y_\psi \Phi \bar{\psi}_L \psi_R + h.c.) \rightarrow (m_\psi e^{i\phi/v_\phi} \bar{\psi}_L \psi_R + h.c.), \quad (2.40)$$

The term $e^{i\phi/v_\phi}$ can then be removed via a field-dependent transformation of ψ ,

$$\psi \rightarrow e^{-i\gamma^5 \frac{\phi}{2v_\phi}} \psi, \quad (2.41)$$

which generates interactions between ϕ , ψ , and the gauge bosons under which ψ is charged. Denoting the $SU(N)$ field strength tensor by $G_{\mu\nu}^a$, and the one for $U(1)$ by $F_{\mu\nu}$, the Lagrangian describing the interactions of ϕ are then,

$$\mathcal{L} \supset -\frac{\phi}{v_\phi} \frac{g^2}{32\pi^2} G_{\mu\nu}^a \tilde{G}^{a\mu\nu} - 2N_c Q^2 \frac{\phi}{v_\phi} \frac{e^2}{32\pi^2} F_{\mu\nu} \tilde{F}^{\mu\nu} + \frac{\partial_\mu \phi}{2v_\phi} \bar{\psi} \gamma^\mu \gamma^5 \psi, \quad (2.42)$$

where e is the gauge coupling corresponding to $U(1)$, and g the one of the $SU(N)$. The coupling between ϕ , $G_{\mu\nu}^a$, $F_{\mu\nu}$, and ψ are then the usual axion–matter interactions as in Eqs. (2.32) and (2.33) with $\alpha_G = g^2/(32\pi^2)$, $\alpha_F = 2N_c Q^2 e^2/(32\pi^2)$, $c_\psi^0 = 1$ and $f = v_\phi$. The only missing ingredient to identify ϕ as an axion is that $g \sim \mathcal{O}(1)$ at some energy scale $\Lambda < f$, so that it attains an effective potential as in Eq. (2.34).

An illuminating thing to note about Eq. (2.42) is that the coupling ϕ to $\bar{\psi} \gamma^\mu \gamma^5 \psi$ can simply be rewritten as $(\partial_\mu \phi / v_\phi) J_{\text{PQ},\psi}^\mu$. Expressed in this way, it is then just the expected coupling of a NG boson of a spontaneously broken symmetry group to the current that is associated with this symmetry. In taking this perspective, one can then deduce that J_{PQ}^μ is in fact anomalous by integrating Eq. (2.42) in parts:

$$\partial_\mu J_{\text{PQ}}^\mu = -\frac{g^2}{16\pi^2} G_{\mu\nu}^a \tilde{G}^{a\mu\nu} - \frac{e^2 N_c Q^2}{16\pi^2} F_{\mu\nu} \tilde{F}^{\mu\nu}, \quad (2.43)$$

Following these statements in reverse, one finds that the coupling of ϕ to $SU(N)$ and $U(1)$ bosons is therefore a consequence of J_{PQ}^μ being anomalous. This is also the reason we needed to charge ψ under $SU(N)$, if we had not, there would be no associated anomaly for J_{PQ}^μ , and we would not generate the vital interaction between ϕ and $G\tilde{G}$ in Eq. (2.42) that is needed for the axion to attain a mass later on.

This observation can be used to construct a variety of possible realizations of the PQ mechanism. The starting point is to associate PQ charges Q_i to a collection of fermions ψ_i , such that the fermionic part of the PQ current is $J_{\text{PQ},\psi}^\mu = \sum_i Q_i \bar{\psi}_i \gamma^\mu \gamma^5 \psi_i$. The success of the PQ mechanism then only requires the gauge charges of the fermions ψ_i to be such that J_{PQ}^μ is anomalous with respect to a confining non-Abelian gauge group. Of course, the details of the Higgs sector can also be altered.

For a reader interested in specific realizations of the PQ mechanism, a good starting point are the classic models for the QCD axion: the *Weinberg–Wilzcek model* (WW) [69, 70], the *Kim–Shifman–Vainshtein–Zakharov model* (KSVZ) [328, 329], and the *Dine–Fischler–Srednicki–Zhitnitsky model*

(DFSZ) [330, 331]. Particularly noteworthy among these is the WW model, the historically first model to realize the PQ mechanism. It is an elegant attempt to achieve the PQ mechanism alongside electroweak-symmetry breaking (EWSB). However, its elegance is its downfall; this construction entails that the axion decay constant f has to be the same as the electroweak vacuum expectation value, i.e., $f \sim 250$ GeV. Consequently, the axion couples strongly to SM matter, and the model was swiftly ruled out, e.g., by beam-dump experiments [304]. The KSVZ and DFSZ model decouple the PQ mechanism from EWSB, leading to much weaker interactions between the QCD axion and the SM.

2.3.2 Axions from string theory

Finally, before proceeding to discuss axions in the context of axion inflation, let us comment about axions that arise from string theory.¹¹ Axions are a generic low-energy prediction of string theory [332], a fact which is also referred to as the *string axiverse* [333]. Axions in string theory come in two types. Closed-string axions, a_i , typically appear together with a scalar modulus field (*saxion*), σ_i , as a consequence of supersymmetry (SUSY). Open-string axions, on the other hand, are more akin to the QFT axions we have discussed so far. They are the angular components of matter fields that break global symmetries.

String theories generically feature additional spacetime dimensions beyond the usual four. These additional dimensions are compactified. For example, for an unbroken SUSY with chiral matter, the compact manifolds are Calabi–Yau [334]. Closed-string axions are the zero-modes of higher-dimensional antisymmetric tensors compactified on the Calabi–Yau manifold [335]. The number of these axions corresponds to the number of topologically distinct ways that the antisymmetric tensor fields wrap around the compact space. The moduli field σ_i parametrize the size of these compact cycles (p -cycle). Integrating the antisymmetric tensors over these cycles gives the axion fields.

The antisymmetric tensor fields from which closed string axions originate are gauge-invariant. This implies that the axion enjoys a shift symmetry, just like in the PQ mechanism. Therefore, they are massless at all orders in perturbation theory. The axion decay constant is determined from the moduli fields, such that, schematically,

$$f_{a,i} \sim \frac{M_{\text{P}}}{\sigma_i} \lesssim M_{\text{P}}. \quad (2.44)$$

The inequality is necessary to ensure that the EFT description of the axion is valid.

Instantons with finite energy S_0 give masses to the axions. Similarly to the field-theory axion, this leads to an effective potential for the axion as in Eq. (2.34). For example, the potential height could be set by $\Lambda^4 \sim m_{\text{SUSY}}^2 M_{\text{P}}^2 e^{-S_0}$, where m_{SUSY} is the scale of SUSY breaking. The gauge-coupling relevant to determine S_0 will generically depend on the moduli field, typically $g^2 \propto 1/\sigma_i$, such that the typical axion mass will scale as

$$m_\phi^2 \sim \frac{m_{\text{SUSY}}^2 M_{\text{P}}^2}{f^2} e^{-\alpha \sigma_i}. \quad (2.45)$$

This scaling highlights that the axion mass is exponentially sensitive to the stabilized value of its modulus. However, if SUSY breaking stabilizes the moduli field, it attains a mass $m_\sigma \sim m_{\text{SUSY}}$, and the axion mass can be viewed as constant at low energies.

¹¹Our summary is based on the account given in Ref. [76].

Chapter 3

Axion inflation

In the previous two chapters, we have discussed two, *a priori* independent, topics; in chapter 1, we have introduced the inflationary paradigm, while in chapter 2, we have discussed axions from the perspective of particle physics and high-energy theory. In this chapter, we connect these two topics by focussing on models of slow-roll inflation in which the inflaton field is identified with an axion field. Such inflationary models can also be simply referred to as models of *axion inflation*.

The appeal of these models is that axions are naturally equipped with an approximate shift symmetry which protects their potential from radiative corrections. Therefore, the identification of the inflaton with an axion is thought to alleviate the usual η -problem of conventional inflation models [59]. Meanwhile, this same shift symmetry still permits couplings between axions and other matter fields, such as fermions and gauge fields. Incorporating these couplings into a model of inflation then entails the direct production of particles during inflation itself, while still ensuring the radiative stability of the inflationary potential [53, 54, 64, 336]. Models of axion inflation are therefore a natural setting in which to explore the effects of inflationary particle-creation processes, for instance their dynamical impact on the inflationary background [55, 64, 337], their imprint on the spectrum of inflationary perturbations [55, 56, 338], or as a mechanism for preheating [78].

The purpose of this chapter is to motivate and establish the concrete model of axion inflation at the center of this thesis. This model is constructed with the particular aim of exploring the physics associated with Abelian gauge-field production during inflation. To this end, we begin in Sec. 3.1 by contextualizing this model within the broader landscape of axion inflation models. In Sec. 3.2, we then establish the mathematical framework underlying our particular model of axion inflation; defining its action, deriving its fundamental equations of motions, and discussing the treatment of Abelian gauge fields in this model. Building on this discussion, we then broadly review two aspects of this particular model that we will further explore in the following chapters of this thesis: in Sec. 3.3, we discuss the dynamical production and backreaction of Abelian gauge fields during axion inflation, while in Sec. 3.4 we highlight their phenomenological implications.

3.1 What is axion inflation?

For a first overview of axion inflation, we should first conceptualize that axion inflation does not constitute a single framework which unifies the idea of slow-roll inflation with axion physics. Rather, the term “axion inflation” can be used to broadly refer to an inflationary model in which the inflaton field (or fields in the case of multifield model) inherits one or more of the characteristic features associated with axions. For instance, this can mean a model of slow-roll inflation with a specific inflationary potential that is motivated by axion physics (i.e., the cosine potential associated with axions), or a model in which the inflaton is simply a pseudoscalar that couples to other fields via one of the typical axion–matter interactions.

In fact, it is useful to conceptually separate axion-motivated slow-roll models from models that include axion–matter interactions, as they are often studied for different reasons in the literature. Axion-motivated slow-roll models are interesting because they can be thought of as inflationary models that avoid the η -problem, while axion–matter interactions offer the potential for exploring the effects of inflationary particle production. Of course, one may be ultimately interested in combining these two aspects of axion inflation, and indeed, studies of this kind exist (see, e.g., Refs [64, 88, 103, 339, 340]), but at present, they are often considered separately from one another.

Therefore, in what follows we also discuss these two aspects separately: first, we review axion-motivated slow-roll models, and then give an overview regarding the effects of particle creation due to axion–matter interactions, with a focus on gauge fields. This separation aligns with the motivation of this thesis to explore the production of Abelian gauge fields during inflation. In particular, in later parts of this thesis, we will not concern ourselves with particular axion-motivated slow-roll scenarios, and instead take an agnostic stance regarding the inflaton’s potential.

3.1.1 Slow-roll with axions

The model of *natural inflation* is the simplest, and historically first, realization of slow-roll inflation with an axion. It was first proposed by FREEZE, FRIEMAN and OLINTO [59] as a single-field slow-roll model which manages to avoid the fine-tuning issues associated with the η -problem. At its core, natural inflation is akin to hilltop models [151], and the inflaton potential takes the typical shape for an axion field, i.e.,

$$V(\phi) = \Lambda^4 \left[1 + \cos \left(\frac{\phi}{f} \right) \right]. \quad (3.1)$$

Inflation thus occurs for ϕ close to the potential maximum. The success of slow-roll inflation only depends on the value of f , requiring $f \gtrsim M_{\text{P}}/6$, while Λ sets the scale of inflation, and the inflaton mass $m_\phi \sim \Lambda^4/f^2$.

At first sight, natural inflation is an extremely promising model of inflation with a particularly clear, yet simple high-energy physics motivation (see also Refs. [341, 342] for similar constructions); the axion field responsible for inflation is thought to obtain its potential from nonperturbative quantum effects, making it radiatively stable. Such an axion can, for instance, originate from a PQ mechanism, or it may have a string-theory origin [327] (see also chapter 2). However, this construction also raises theoretical concerns, as successful inflation on the potential in Eq. (3.1) generally seems to necessitate a large, nearly Planckian, axion decay constant, $f \sim M_{\text{P}}$. In particular, it is questionable if the PQ mechanism is indeed a plausible explanation for an axion field with such a large decay constant. The PQ mechanism assumes the spontaneous breaking of a global symmetry at $f \sim M_{\text{P}}$, but these symmetries are generally believed to be already broken by quantum-gravity effects at such scales (see e.g., Refs. [343, 344]). Meanwhile, a string-theory origin of such an axion field may be more viable, although specific realizations with $f \sim M_{\text{P}}$ are difficult to achieve [345, 346].

From a modern perspective, the argument against natural inflation is instead far simpler; the model is ruled out by current CMB observations as it predicts either too low a scalar spectral tilt, n_s , or too large a value of the tensor-to-scalar ratio r (see, e.g., Ref. [347] or cf. Fig. 1.5).

In light of these problems, a variety of other axion-motivated inflation scenarios have been proposed which can both avoid the need for a Planckian decay constant and can be in agreement with CMB constraints. The first is a class of string-theory-motivated models called *axion monodromy inflation* [62, 63]. Therein, the axion potential is not periodic like in Eq. (3.1), but rather decreases

as the axion field completes one revolution in field space going from zero to $2\pi f$;

$$V(\phi) = \mu^{4-p} \phi^p + \Lambda^4 \cos\left(\frac{\phi}{f}\right). \quad (3.2)$$

Here, μ is an energy scale of mass dimension one, and p is a model-dependent exponent (typically, $p < 2$ [348]). In particular, models of monodromy inflation with a flattened potential can be in agreement with CMB constraints [349], while the cosine modulations of the potential can lead to distinct features in the scalar power spectrum [350]. Recently, it was also noted that the cosine modulations in this potential can induce large scalar anisotropies associated with primordial-black-hole (PBH) formation [351].

Another way to circumvent the issues of natural inflation is to consider multifield models of inflation. The first such example is *aligned axion inflation* [60, 88, 352, 353]. The model proposes two axion fields, both in a potential as in Eq. (3.1). The inflationary trajectory is then a linear combination of the two axion fields corresponding to a flat direction in their scalar potential. Another notable model is *N-flation* [61]. In the spirit of assisted inflation [154], N-flation proposes N axions whose combined Hubble drag supports inflation, even when each individual axion potential is too steep for the standard natural-inflation trajectory. The model has a rich CMB phenomenology [354], including the possibility for detectable levels of scalar non-Gaussianity [355, 356]. Generically, the large number of axions needed for N-flation implies a string-theory origin (see, e.g., Refs. [357, 358]).

3.1.2 Axion–gauge-field interactions during inflation

Let us next turn our attention to the axion–gauge field interactions in axion inflation:¹

$$\mathcal{L}_{\text{int}} = -\frac{\beta\phi}{4M_{\text{P}}} F_{\mu\nu}^a \tilde{F}^{a\mu\nu}. \quad (3.3)$$

Here, ϕ is the pseudoscalar inflaton, $F_{\mu\nu}^a$ is the field-strength tensor for either an Abelian or a non-Abelian gauge field, and, $\tilde{F}^{a\mu\nu}$ is its dual (cf. Eq. (2.7)). Moreover, β denotes a dimensionless axion–vector coupling. Below, we first give an overview regarding the dynamical and phenomenological implications of this interaction from a general perspective, without distinguishing between non-Abelian and Abelian gauge fields. We then briefly highlight the main differences between these two kinds of fields, before we then focus entirely, and in more detail, on the Abelian axion–gauge-field interaction in the subsequent section, and this thesis as a whole.

The inclusion of the interaction in Eq. (3.3) into an inflation model has two important dynamical consequences: the inflaton field directly sources gauge fields, but these fields may also backreact on its dynamics. The creation of gauge fields proceeds via a tachyonic instability which is triggered by the continuous rolling of the inflaton field during inflation. Because of the pseudoscalar nature of the axion, this production mechanism sources helical gauge fields [53, 360–362]. Meanwhile, the homogeneous inflaton background field, $\varphi = \langle\phi\rangle$, is subjected to a novel friction term [55, 64],

$$\ddot{\varphi} + 3H\dot{\varphi} + V_{,\phi}(\varphi) = -\frac{\beta}{4M_{\text{P}}} \langle F_{\mu\nu}^a \tilde{F}^{a\mu\nu} \rangle. \quad (3.4)$$

If the inflaton field sources enough gauge-field friction, the right-hand side term can become dynamically important and may cause a breakdown of the standard slow-roll attractor.

¹The fermionic interaction is also studied, albeit less frequently, see for instance Refs. [336, 337, 359].

Besides their dynamical effects, these gauge fields can also source scalar and tensor perturbation [55, 56, 58, 338, 361], reheat the Universe after inflation [78], source large-scale magnetic fields that may have survived until the present-day Universe [53, 54, 94], or supply a necessary source of CP violation for baryogenesis [101, 102].

In the context of axion inflation, the important difference between Abelian and non-Abelian fields is that an Abelian gauge field encompasses only vectorial degrees of freedom, while non-Abelian gauge fields are also composed of scalar and tensor modes. This affects how Abelian and non-Abelian fields can backreact on the inflationary dynamics and how they can source inflationary perturbations. The difference in the sourcing of inflationary perturbations is easy to understand: A non-Abelian gauge field can source scalar and tensor modes linearly [338, 361], while the Abelian gauge field can only source scalar and tensor modes via a source term that is quadratic in the field's vectorial modes [55, 56, 58]. More profound is the effect this has on backreaction, which we shall briefly elaborate on in what follows.

In the Abelian case, the vector-like gauge field cannot assume a homogeneous and isotropic configuration, as the inflaton field only sources vectorial electric and magnetic fields, $\mathbf{E}$ and $\mathbf{B}$.² The gauge field can then only backreact onto the inflaton dynamics in Eq. (3.4) via a scalar combination of $\mathbf{E}$ and $\mathbf{B}$, namely

$$\langle F_{\mu\nu} \tilde{F}^{\mu\nu} \rangle = -4 \langle \mathbf{E} \cdot \mathbf{B} \rangle . \quad (3.5)$$

The chiral production mechanism via the rolling pseudoscalar field ensures that $\langle \mathbf{E} \cdot \mathbf{B} \rangle$ is non-zero, so inflation can also be supported away from the standard slow-roll attractor [55]. However, the Abelian gauge field does not provide a stable source of friction, and once it becomes important, the dynamics of inflation become highly nonlinear [107–109].

Meanwhile, a non-Abelian gauge field can form an isotropic and homogeneous configuration, namely [363, 364]

$$A_0^a = 0 , \quad A_i^a(t) = \delta_i^a Q(t) , \quad (3.6)$$

where A_μ^a is the gauge field and $Q(t)$ a time-dependent background vacuum expectation value. The model of *chromonatural inflation* [64] corresponds to the case where Q balances the potential gradient of the natural inflation potential.³ Before the isotropic gauge-field configuration in Eq. (3.6) forms, the non-Abelian field acts as multiple copies of the Abelian field, and the isotropic configuration in Eq. (3.6) can naturally emerge through gauge-field production [362, 371]. Once it has formed, there is a regime in which its friction is stable [360].

This difference between non-Abelian and Abelian gauge fields draws a dividing line between models that incorporate the non-Abelian axion–vector interaction, and those constructed around the Abelian one, as both types of interactions are faced with their own unique challenges and perspectives. In the literature, one sometimes draws a difference between models incorporating the former interaction as *non-Abelian axion inflation*, and those built on the latter as *Abelian axion inflation*. In particular, our focus in this thesis will be on the Abelian axion inflation model, which we shall now turn to.

²It is common practice to refer to the Abelian gauge field in axion inflation as an electromagnetic field and its corresponding three-vector fields as electric and magnetic fields. This is done in analogy to the SM photon field, but does not mean that only this case is considered in the literature. We also adopt this nomenclature throughout this thesis.

³Note that some authors also study the non-Abelian axion–gauge-field interaction in the context of warm inflation [365–369], including an intermediate regime between chromonatural and warm inflation [370].

3.2 The model of Abelian axion inflation

In this section, we establish the mathematical formulation of Abelian axion inflation which we shall rely on as a foundation for the later chapters of this thesis. In particular, we shall introduce the set of equations that describe the inflationary dynamics of this model, including both the production of gauge fields via the inflaton field and their subsequent backreaction.

Historically, the model we present here has been developed in two stages. In their seminal work, GARRETSON, FIELD and CARROLL established the basic mechanism of Abelian gauge-field production via the axion–vector interaction in Eq. (3.3) to propose it as a mechanism for magnetogenesis [53]. This idea was revisited by ANBER and SORBO who presented the concrete model of axion inflation in the form as it is studied today [54, 55].

Seeing as this is the central model of axion inflation that we study in this thesis, we will refer to it simply as “axion inflation”, we implicitly assume that the gauge fields we are discussing are Abelian, and we only emphasize their Abelian nature if the context requires it.

3.2.1 The equations of motion

We begin by defining the action for our model. This action describes the pseudoscalar inflaton field, ϕ and its interaction with an Abelian gauge field A_μ via the Chern–Simons operator in Eq. (3.3):

$$S = \int d^4x \sqrt{-g} \left[\frac{1}{2} g^{\mu\nu} \partial_\mu \phi \partial_\nu \phi - V(\phi) - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} - \frac{\beta \phi}{4M_P} F_{\mu\nu} \tilde{F}^{\mu\nu} \right]. \quad (3.7)$$

As before, $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is the field-strength tensor, and $\tilde{F}^{\mu\nu} = \varepsilon^{\mu\nu\alpha\beta} F_{\alpha\beta} / (2\sqrt{-g})$ is its dual, and β is a dimensionless axion–vector coupling. The inflaton potential, $V(\phi)$, is at this point left arbitrary, and we merely assume that it can support a stage of slow-roll inflation.

Varying this action with respect to ϕ and A_μ yields the Euler–Lagrange equations for this system, which, when combined with the Bianchi identity for $\tilde{F}_{\mu\nu}$, gives,

$$\frac{1}{\sqrt{-g}} \partial_\alpha [\sqrt{-g} g^{\alpha\mu} \partial_\mu \phi] + V_{,\phi}(\phi) + \frac{\beta}{4M_P} F_{\mu\nu} \tilde{F}^{\mu\nu} = 0, \quad (3.8a)$$

$$\frac{1}{\sqrt{-g}} \partial_\alpha [\sqrt{-g} g^{\mu\alpha} g^{\nu\beta} F_{\mu\nu}] + \frac{\beta}{M_P} \tilde{F}^{\alpha\beta} \partial_\alpha \phi = 0, \quad (3.8b)$$

$$\frac{1}{\sqrt{-g}} \partial_\alpha [\sqrt{-g} \tilde{F}^{\alpha\beta}] = 0. \quad (3.8c)$$

As usual, the dynamics of spacetime are governed by the Einstein equations (1.3). The energy–momentum tensor which sources it is obtained by varying Eq. (3.7) with respect to $g^{\mu\nu}$:

$$T_{\mu\nu} = \partial_\mu \phi \partial_\nu \phi - g^{\alpha\beta} F_{\mu\alpha} F_{\nu\beta} - g_{\mu\nu} \left[\frac{1}{2} \partial_\alpha \phi \partial^\alpha \phi - V(\phi) - \frac{1}{4} F_{\alpha\beta} F^{\alpha\beta} \right]. \quad (3.9)$$

Our interest is in inflationary dynamics, so we assume FLRW spacetime with the metric given by Eq. (1.23). Since this spacetime is not invariant under the full Lorentz group, it proves useful to work with electric and magnetic three-vector fields, $\mathbf{E} = (E^1, E^2, E^3)$ and $\mathbf{B} = (B^1, B^2, B^3)$. Invoking gauge freedom to work in temporal gauge, $A_\mu = (0, -\mathbf{A})$, we then define $\mathbf{E}$ and $\mathbf{B}$ as,

$$\mathbf{E} = -\frac{1}{a} \partial_0 \mathbf{A}, \quad \mathbf{B} = \frac{1}{a^2} \text{rot} \mathbf{A}, \quad (3.10a)$$

$$F_{0i} = aE^i, \quad F_{ij} = -a^2 \varepsilon_{ijk} B^k, \quad \tilde{F}_{0i} = aB^i, \quad \tilde{F}_{ij} = a^2 \varepsilon_{ijk} E^k. \quad (3.10b)$$

With this definition, $\mathbf{E}$ and $\mathbf{B}$ are the physical fields measured by a comoving observer.

Just as in the case of the slow-roll inflation in Sec. 1.2.1, one can analyze the system in Eq. (3.8) for the case where the inflaton field separates into a homogeneous background field and a perturbation, $\phi(t, \mathbf{x}) = \varphi(t) + \delta\phi(t, \mathbf{x})$, assuming that the inhomogeneities, $\delta\phi$, are dynamically negligible. The equation of motion (EOM) for φ is then given by

$$\ddot{\varphi} + 3H\dot{\varphi} + V_{,\phi}(\varphi) = \frac{\beta}{M_{\text{P}}} \langle \mathbf{E} \cdot \mathbf{B} \rangle . \quad (3.11)$$

This is simply the homogeneous Klein–Gordon equation for the background inflaton field as in Eq. (1.22), but now with the appearance of an additional friction term due to the electric and magnetic field, as anticipated in Eq. (3.4). To derive Eq. (3.11), we used $\varphi \equiv \langle \phi \rangle$ to separate out the homogeneous part of Eq. (3.8a), with $\langle \cdot \rangle$ denoting the expectation value during inflation. As we see later on, this expectation value can be interpreted in the same way for the gauge-field operators $\mathbf{E}$ and $\mathbf{B}$ as for a scalar field (cf. Sec. 1.3).

The dynamical evolution of the electromagnetic fields is governed by,

$$\text{div } \mathbf{E} = 0, \quad \text{div } \mathbf{B} = 0, \quad (3.12a)$$

$$\dot{\mathbf{E}} + 2H\mathbf{E} - \frac{1}{a} \text{rot } \mathbf{B} + \frac{\beta}{M_{\text{P}}} \dot{\varphi} \mathbf{B} = 0, \quad (3.12b)$$

$$\dot{\mathbf{B}} + 2H\mathbf{B} + \text{rot } \mathbf{E} = 0, \quad (3.12c)$$

which follow from Eqs. (3.8b) and (3.8c) for an FLRW metric. These are simply Maxwell's equations in an expanding spacetime with an additional source term depending on the inflaton velocity, $(\beta/M_{\text{P}})\dot{\varphi}\mathbf{B}$. As we see below, this source is responsible for gauge-field production, and we may already note that the inflaton velocity plays an important role in this production mechanism.

The final equation of importance is the Friedmann equation. In the model we discuss here, it takes the form

$$H^2 = \frac{1}{3M_{\text{P}}^2} \left[\frac{1}{2} \dot{\varphi}^2 + V(\varphi) + \frac{1}{2} \langle \mathbf{E}^2 + \mathbf{B}^2 \rangle \right]. \quad (3.13)$$

The energy density on the right-hand side is simply composed of the inflaton's energy density (cf. Eq. (1.24a)) and the averaged gauge-field energy density, $\rho_{\text{EM}} = \langle \mathbf{E}^2 + \mathbf{B}^2 \rangle / 2$. From the second Friedmann equation, Eq. (1.5b), one can also determine that the condition for accelerated expansion in this model is

$$V(\varphi) \geq \dot{\varphi}^2 + \rho_{\text{EM}}. \quad (3.14)$$

Thus, inflation is ensured if the inflaton's potential dominates the system's total energy density.

In combination, Eqs. (3.11) to (3.13) govern the entire dynamics of axion inflation. Before we can however discuss these dynamics in detail, we first need to understand the mechanism which leads to the production of gauge fields in this model, and which sources the classical background of electromagnetic fields described by $\langle \mathbf{E} \cdot \mathbf{B} \rangle$, $\langle \mathbf{E}^2 \rangle$ and $\langle \mathbf{B}^2 \rangle$.

3.2.2 Gauge-field production and the tachyonic instability

To discuss gauge-field production during axion inflation, we treat the gauge field $\mathbf{A}$ as a quantum field whose dynamical modes classicalize during inflation, similar to our discussion of quantum scalar fields during inflation in Sec. 1.3.

We begin by considering the quantization of $\mathbf{A}$. To this end, we can first note that $\mathbf{A}$ only describes two propagating degrees of freedom in our model, because of the constraint $\text{div } \mathbf{E} = 0$ in Eq. (3.12a). In temporal gauge, this constraint enforces the additional Coulomb condition,

$\text{div } \mathbf{A} = 0$. Therefore, $\mathbf{A}$ only contains two transverse polarization states, which, in our model, are best described in a helicity basis. Following this observation, we then promote $\mathbf{A}$ to an operator, and write it in momentum space as

$$\mathbf{A}(t, \mathbf{x}) = \int \frac{d^3 \mathbf{k}}{(2\pi)^{3/2}} \sum_{\lambda=\pm 1} \left[\epsilon^\lambda(\mathbf{k}) \hat{a}_{\mathbf{k},\lambda} A_\lambda(t, k) e^{i\mathbf{k}\cdot\mathbf{x}} + \epsilon^{\lambda*}(\mathbf{k}) \hat{a}_{\mathbf{k},\lambda}^\dagger A_\lambda^*(t, k) e^{-i\mathbf{k}\cdot\mathbf{x}} \right]. \quad (3.15)$$

Here, $A_\lambda(t, k)$ are mode functions, $\epsilon^\lambda(\mathbf{k})$ are polarization three-vectors, and $\hat{a}_{\mathbf{k},\lambda}$ ($\hat{a}_{\mathbf{k},\lambda}^\dagger$) are annihilation (creation) operators, each with momentum $\mathbf{k}$ and circular polarization $\lambda = \pm 1$, where $k = |\mathbf{k}|$. The circular polarization vectors $\epsilon^\lambda(\mathbf{k})$ are defined through the properties

$$\mathbf{k} \cdot \epsilon^\lambda(\mathbf{k}) = 0, \quad [i\mathbf{k} \times \epsilon^\lambda(\mathbf{k})] = \lambda k \epsilon^\lambda(\mathbf{k}), \quad \epsilon^{\lambda*}(\mathbf{k}) \cdot \epsilon^{\lambda'}(\mathbf{k}) = \delta^{\lambda\lambda'}, \quad (3.16)$$

with $\epsilon_\lambda^*(\mathbf{k}) = \epsilon_{-\lambda}(\mathbf{k}) = \epsilon_\lambda(-\mathbf{k})$. Moreover, we impose canonical commutation relations for the creation and annihilation operators:

$$[\hat{a}_{\mathbf{k},\lambda}, \hat{a}_{\mathbf{k}',\lambda'}^\dagger] = \delta_{\lambda\lambda'} \delta^{(3)}(\mathbf{k} - \mathbf{k}'). \quad (3.17)$$

The mode equation for $A_\lambda(t, k)$ then follows from Eq. (3.12b), and reads

$$\ddot{A}_\lambda(t, k) + H \dot{A}_\lambda(t, k) + \left[\frac{k^2}{a^2} - \lambda \frac{k}{a} \frac{\beta}{M_{\text{P}}} \dot{\varphi} \right] A_\lambda(t, k) = 0. \quad (3.18)$$

Immediately, we should take note that the mode equation (3.18) affects gauge-field modes differently depending on their polarization, λ . In what follows, we now want to study the evolution of the gauge-field modes $A_\lambda(\eta, k)$ for a fixed wavenumber k and for its two polarizations $\lambda = \pm 1$. Firstly, we do this to identify a unique vacuum state for these modes, i.e., we want to impose boundary conditions that uniquely define the creation and annihilation operators in Eq. (3.15), just as we have discussed it for a quantum scalar field in Sec. 1.3. Moreover, this analysis also reveals the mechanism of gauge-field production in our model.

Thus, let us assume an inflationary spacetime with a growing Hubble scale aH , such that Eq. (3.18) can be conveniently rewritten in conformal time, η , as

$$A_\lambda''(\eta, k) + [k^2 - 2\lambda k \xi aH] A_\lambda(\eta, k) = 0. \quad (3.19)$$

Here, for later convenience, we have introduced the variable

$$\xi \equiv \frac{\beta}{M_{\text{P}}} \frac{\dot{\varphi}}{2H}, \quad (3.20)$$

which is called the *instability parameter* for reasons that shall become evident shortly.

To discuss the evolution of gauge-field modes, it proves convenient for the moment to assume that the parameter ξ is constant. This also reflects an *a priori* sensible assumption from the perspective of slow-roll inflation, where both the Hubble rate, H , and the inflaton velocity, $\dot{\varphi}$, vary slowly (cf. Sec. 1.2). As we will see later on, this assumption is not always justified for the model of axion inflation we consider here, but it is an instructive for a first view into the mode of gauge-field production in this model.

Let us then consider a specific gauge-field mode $A_\lambda(\eta, k)$ while its wavelength is much shorter than the Hubble rate, i.e., $k \gg 2|\xi|aH$. In this case, we can see that Eq. (3.19) reduces to the form of a simple harmonic oscillator with a frequency k^2 . Thus, we can impose Bunch–Davies boundary conditions for sub-horizon gauge-field modes, similar to how we have defined them for

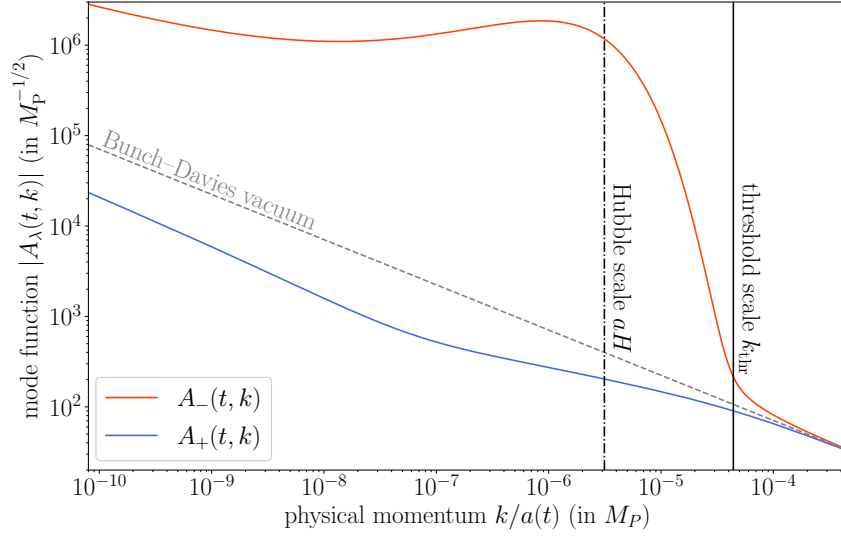


FIGURE 3.1: An example spectrum of gauge-field modes $A_\lambda(t, k)$ illustrating the tachyonic enhancement of a preferred polarization mode. The spectrum is computed for a nearly constant, negative instability parameter ξ . One helicity mode ($\lambda = -1$, red), is enhanced out of its vacuum-like state for modes below the threshold scale in Eq. (3.22) (black, solid), while the other helicity, $\lambda = 1$ (blue), is mildly damped. For reference, we also indicate the amplitude of the Bunch–Davies vacuum (gray, dashed), and the Hubble scale (black, dash-dotted).

a quantum scalar field in Sec. 1.3:

$$A_\lambda(\eta, k) \rightarrow \frac{1}{\sqrt{2k}} e^{-ik\eta}, \quad -k\eta \rightarrow \infty. \quad (3.21)$$

These boundary conditions then uniquely define a vacuum state for the modes $A_\lambda(\eta, k)$ for both polarizations, namely, the Bunch–Davies vacuum.

Next, let us study how the evolution of a gauge-field mode proceeds starting from its vacuum configuration. During inflation, the scale factor a increases exponentially, meaning that the wavenumber k of a mode $A_\lambda(\eta, k)$ may eventually fall below the threshold

$$k \leq k_{\text{thr}} \equiv 2|\xi|aH. \quad (3.22)$$

This threshold scale is linked to a remarkable feature in the mode equation (3.19): its tachyonic instability. For a gauge-field mode with $k \leq k_{\text{thr}}$ and a polarization $\lambda = \text{sign}(\xi)$ the frequency term in Eq. (3.19) becomes negative. As a result, the mode is tachyonically amplified, and stops oscillating. Meanwhile, this does not happen for the mode with the opposite polarization, $\lambda = -\text{sign}(\xi)$, because k is strictly positive, and so its frequency term in Eq. (3.19) cannot become negative. Correspondingly, this mode therefore is not amplified and remains in an oscillator-like configuration close to the Bunch–Davies vacuum.

For a visualization of this tachyonic enhancement, in Fig. 3.1, we show the spectrum of gauge-mode functions $A_\lambda(\eta, k)$ for a nearly constant and negative value of ξ . We can clearly recognize the enhancement of the negative (red) polarization mode above its Bunch–Davies vacuum state (dashed gray) below the threshold scale k_{thr} (black solid). Meanwhile, the positive helicity mode (blue) is suppressed with respect to the negative one, and its amplitude even falls mildly below that of the Bunch–Davies vacuum.

The dependence of the threshold scale in Eq. (3.22) on the parameter ξ also explains why it

is referred to as the instability parameter; it regulates which wavenumbers k have experienced the tachyonic instability at a given moment of time, and thus, which modes are amplified out of their vacuum state. In particular, larger values of ξ correspond to a larger fraction of amplified modes. Meanwhile, the sign of ξ regulates which helicity mode is amplified.

This amplification of the gauge-field modes is then the mechanism of gauge-field production in axion inflation. Therefore, the parameter ξ is also referred to as the *gauge-field-production parameter*.⁴ We can then connect this mechanism back to physical intuition by noting that $\xi \propto \dot{\varphi}$. From this, we can recognize that it is indeed the rolling of the inflaton field which sources gauge fields in this model, as it cannot take place when the inflaton's velocity is zero. In particular, during slow-roll inflation, the inflaton velocity is non-vanishing and is traversing into one preferred direction down its potential slope (i.e., $\dot{\varphi}$ selects a sign). Since the inflaton is a pseudoscalar, this selection of a preferred direction of motion spontaneously breaks parity, which then explains why we find a helical mechanism for producing gauge fields in axion inflation.

Now that we have understood this mechanism, we should then return to the mathematical formulation of our axion inflation model to see how this gauge-field production feeds into the gauge-field expectation values appearing in its equations of motion. More specifically, we want to motivate explicit expressions for the expectation values $\langle \mathbf{E} \cdot \mathbf{B} \rangle$ and $\langle \mathbf{E}^2 + \mathbf{B}^2 \rangle$ appearing, respectively, as friction terms in the Klein–Gordon equation (3.11), and as an energy-density contribution in the Friedmann equation (3.13). This discussion then allows us to write the equations of motions in a closed form, thus setting the stage for discussing axion inflation from a fully dynamical perspective.

3.2.3 Gauge-field backreaction

The computation of the expectation values $\langle \mathbf{E} \cdot \mathbf{B} \rangle$ and $\langle \mathbf{E}^2 + \mathbf{B}^2 \rangle$ is, in principle, straightforward, as the selection of Bunch–Davies boundary conditions uniquely defines a vacuum state that is annihilated by the operator $\hat{a}_{\mathbf{k},\lambda}$. We can therefore interpret $\langle \cdot \rangle$ as a quantum expectation value with respect to this vacuum state, and we can evaluate the quantities $\langle \mathbf{E} \cdot \mathbf{B} \rangle$ and $\langle \mathbf{E}^2 + \mathbf{B}^2 \rangle$ by expressing $\mathbf{E}$ and $\mathbf{B}$ in terms of gauge-field mode functions, polarization vectors and ladder operators using Eqs. (3.10a) and (3.15). By applying Wick's theorem, one readily obtains integral expressions for $\langle \mathbf{E} \cdot \mathbf{B} \rangle$ and $\langle \mathbf{E}^2 + \mathbf{B}^2 \rangle$:

$$\langle \mathbf{E}^2 + \mathbf{B}^2 \rangle = \int \frac{dk}{k} \frac{k^3}{2\pi^2 a^2} \sum_{\lambda} \left(|\dot{A}_{\lambda}(t, k)|^2 + \frac{k^2}{a^2} |A_{\lambda}(t, k)|^2 \right), \quad (3.23a)$$

$$\langle \mathbf{E} \cdot \mathbf{B} \rangle = \int \frac{dk}{k} \frac{k^4}{2\pi^2 a^3} \sum_{\lambda} \lambda \operatorname{Re}[\dot{A}_{\lambda}(t, k) A_{\lambda}^*(t, k)]. \quad (3.23b)$$

A priori, these integrals over the wavenumber k of the mode functions $A_{\lambda}(\eta, k)$ should run from zero to infinity. However, we know from our discussion of the quantum scalar field in Sec. 1.3 that such integrals are likely UV-divergent, and indeed, one finds this to be true for $\langle \mathbf{E} \cdot \mathbf{B} \rangle$ and $\langle \mathbf{E}^2 + \mathbf{B}^2 \rangle$: at a given moment of time t , the naive expressions for $\langle \mathbf{E} \cdot \mathbf{B} \rangle$ and $\langle \mathbf{E}^2 + \mathbf{B}^2 \rangle$ sum over infinitely many modes which are still in a near vacuum-like state deep inside the Hubble horizon $k \gg aH$. Thus, to properly define expressions for $\langle \mathbf{E} \cdot \mathbf{B} \rangle$ and $\langle \mathbf{E}^2 + \mathbf{B}^2 \rangle$, we need to regularize this UV divergence. Just as discussed in Sec. 1.3, the simplest way to do so is to identify a UV regulator scale which separates vacuum modes from modes that are no longer in a vacuum-like state at any given moment of time, t .

⁴Referring to ξ as a parameter is a historical convention, as the early literature assumed scenarios in which ξ is constant. In a dynamical setting, which we will be mainly interested in for this thesis, ξ is time-dependent as $\dot{\varphi}$ and H both change. Regardless, we shall still refer to ξ as a parameter in this thesis.

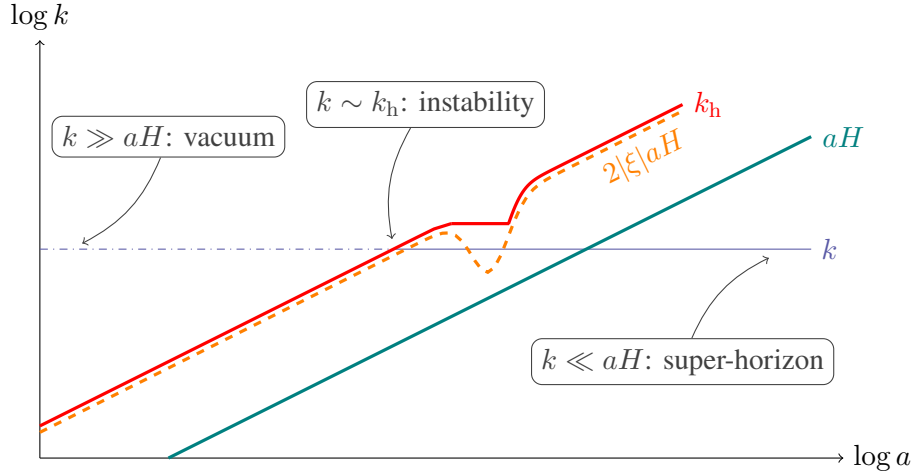


FIGURE 3.2: The evolution of a gauge-field mode, $A_\lambda(t, k)$, for a fixed wavenumber k which undergoes a tachyonic enhancement (blue). Initially, the mode is in a vacuum-like state ($k \gtrsim k_h$); then, it reaches the instability scale, $k \sim k_h$ and is amplified; finally, the mode exits the horizon once $k \sim aH$. The change in $A_\lambda(t, k)$ from near vacuum-like to amplified is indicated by the change from the dash-dotted to the solid line. Also indicated is the difference between k_h (red) and $2|\xi|aH$ (orange). Although k crosses $2|\xi|aH$ a second time, $A_\lambda(t, k)$ is still enhanced and therefore not in a vacuum-like state.

A first naive guess for such a scale is the threshold scale k_{thr} in Eq. (3.22), which separates modes that have been tachyonically enhanced $k \lesssim k_{\text{thr}}$ from modes that are still in a near Bunch–Davies state $k \gtrsim k_{\text{thr}}$. Indeed, we can recognize from Fig. 3.2 that k_{thr} performs this separation well; the amplitude of modes with $k \gtrsim k_{\text{thr}}$ are near that of the Bunch–Davies vacuum, while modes with $k \lesssim k_{\text{thr}}$ are not.⁵ One issue with this identification is that $k_{\text{thr}} = 2|\xi|aH$ is only a suitable separator between vacuum and enhanced modes as long as $|\xi|$ changes adiabatically slowly. As we see later on, $|\xi|$ can, in fact, undergo rapid changes in certain dynamical regimes. In this case, a mode of wavenumber k can briefly become larger than k_{thr} at a certain time t , even though it has already underwent a tachyonic enhancement at an earlier time, $t' \leq t$, and has thus already been amplified out of its vacuum-like state. Thus, we require an improved regularization scale which ensures that, at a given moment of time t , all modes below this scale have undergone a tachyonic enhancement, while all modes above it have not.

The appropriate scale which performs this task is the *instability scale*, k_h , which we can simply define as [128]

$$k_h(t) \equiv \max_{t' \leq t} \{2a(t')H(t')|\xi(t')|\} . \quad (3.24)$$

In an inflationary system where aH increases monotonically and $|\xi|$ only changes slowly, this reproduces our naive expectation, $k_h = k_{\text{thr}}$. Meanwhile, this definition ensures monotonic growth even if $|\xi|$ changes rapidly, so that any mode which *ever* experienced the tachyonic instability has a wavenumber $k \leq k_h$, while any vacuum-like mode has $k \geq k_h$. With this definition, a mode $A_\lambda(t, k)$ with $\lambda = \text{sign}(\xi)$ is first in a Bunch–Davies-like state for $k \gtrsim k_h$, then is tachyonically enhanced when $k \sim k_h$, and then exits the horizon for $k \lesssim aH$. The fact that k_h grows monotonically ensures that the transition between classical and vacuum modes is clearly defined, irrespective of how ξ evolves. An illustration of this logic, including the different scales and their interpretation is given in Fig. 3.2.

⁵Of course, this only works for the amplified polarization mode, $A_\lambda(\eta, k)$, but their contribution to $\langle \mathbf{E} \cdot \mathbf{B} \rangle$ and $\langle \mathbf{E}^2 + \mathbf{B}^2 \rangle$ will dominate seeing as they are enhanced, see also footnote 6.

This instability scale is then a suitable regulator for the expectation values $\langle \mathbf{E} \cdot \mathbf{B} \rangle$ and $\langle \mathbf{E}^2 + \mathbf{B}^2 \rangle$, and we can finally properly define these quantities via the integrals⁶

$$\langle \mathbf{E}^2 + \mathbf{B}^2 \rangle = \int_0^{k_h(t)} \frac{dk}{k} \frac{k^3}{2\pi^2 a^2} \sum_{\lambda} \left(|\dot{A}_{\lambda}(t, k)|^2 + \frac{k^2}{a^2} |A_{\lambda}(t, k)|^2 \right), \quad (3.25a)$$

$$\langle \mathbf{E} \cdot \mathbf{B} \rangle = \int_0^{k_h(t)} \frac{dk}{k} \frac{k^4}{2\pi^2 a^3} \sum_{\lambda} \lambda \text{Re}[\dot{A}_{\lambda}(t, k) A_{\lambda}^*(t, k)], \quad (3.25b)$$

which follow from Eqs. (3.10a), (3.15), and the application of Wick's theorem. One may convince oneself that these integrals indeed diverge in the limit $k \rightarrow \infty$ by using Eq. (3.21).

From these expressions, we can also gain an intuitive understanding on how gauge-field production sources $\langle \mathbf{E} \cdot \mathbf{B} \rangle$ and $\langle \mathbf{E}^2 + \mathbf{B}^2 \rangle$: the rolling inflaton field continuously amplifies modes out of their vacuum state, which then contribute to classical large-scale electromagnetic field configurations for wavelengths larger than $1/k_h$.

Finally, we may then express the entire system of dynamical equations for axion inflation as:

$$\ddot{\varphi} + 3H\dot{\varphi} + V_{,\phi}(\varphi) = \frac{\beta}{M_{\text{P}}} \int_0^{k_h(t)} \frac{dk}{k} \frac{k^4}{2\pi^2 a^3} \sum_{\lambda} \lambda \text{Re}[\dot{A}_{\lambda}(t, k) A_{\lambda}^*(t, k)]. \quad (3.26a)$$

$$H^2 = \frac{1}{3M_{\text{P}}^2} \left[\frac{1}{2} \dot{\varphi}^2 + V(\varphi) + \frac{1}{2} \int_0^{k_h(t)} \frac{dk}{k} \frac{k^3}{2\pi^2 a^2} \sum_{\lambda} \left(|\dot{A}_{\lambda}(t, k)|^2 + \frac{k^2}{a^2} |A_{\lambda}(t, k)|^2 \right) \right], \quad (3.26b)$$

$$\ddot{A}_{\lambda}(t, k) + H\dot{A}_{\lambda}(t, k) + \left[\frac{k^2}{a^2} - \lambda \frac{k}{a} \frac{\beta}{M_{\text{P}}} \dot{\varphi} \right] A_{\lambda}(t, k) = 0. \quad (3.26c)$$

These follow from the Klein–Gordon and Friedmann equations in Eqs. (3.11) and (3.13), and are coupled to the mode equation, Eq. (3.19). Evidently, these equations are a highly nonlinear set of coupled integro-differential equations, which one needs to solve to fully study the dynamical interplay between gauge-field production and backreaction in axion inflation.

3.3 The dynamics of Abelian axion inflation

Having established the fundamental set of equations describing the evolution of the inflaton field, the creation of gauge fields, and their backreaction, we now want to consider these processes from a dynamical perspective. To this end, we give a broad overview on how one can solve the equations of motions in Eq. (3.26), and discuss basic features of the resulting dynamics.

First, we need to recognize that a fully general discussion of the inflationary dynamics in axion inflation is not feasible because of the nonlinear structure of its underlying equations of motions. However, one can identify two distinct dynamical regimes in this model with varying degrees of

⁶The introduction of this UV cutoff is the most commonly adopted regularization scheme in the literature on axion inflation (see, e.g., Refs. [83, 103, 109, 128, 372]). This scheme is, however, mainly suitable for a strong tachyonic enhancement ($|\xi| \gtrsim 3$) where the separation between vacuum and enhanced modes is pronounced, and the integrands in Eq. (3.25) are dominated by the enhanced polarization mode [103]. As an alternative, the authors of Refs. [373, 374], regularized $\langle \mathbf{E} \cdot \mathbf{B} \rangle$ and $\langle \mathbf{E}^2 + \mathbf{B}^2 \rangle$ using adiabatic subtraction, albeit only in situations where gauge-field production can be studied analytically (weak backreaction, see Sec. 3.3). In this thesis, we mainly study gauge-field production in situations that require numerical techniques, so we rely entirely on the practically simple UV cutoff as a regulator.

complexity regarding their description. In the first, the dynamical impact of gauge-field production is small, the terms $\langle \mathbf{E} \cdot \mathbf{B} \rangle$ and $\langle \mathbf{E}^2 + \mathbf{B}^2 \rangle$ is negligible, and the mode equation (3.18) does not backreact on the Klein–Gordon and Friedmann equations. We call this the *weak-backreaction regime*. In the opposite case, gauge-field backreaction affects the dynamics of inflation itself, and we are forced to treat the full nonlinear EOMs in Eq. (3.26) as one. We refer to this second case as the *strong-backreaction regime*.

To quantify the difference between these two regimes, it proves useful to introduce two *back-reaction parameters* [103],

$$\delta_{\text{KG}} \equiv \frac{\beta}{M_{\text{P}}} \frac{|\langle \mathbf{E} \cdot \mathbf{B} \rangle|}{|3H\dot{\phi}|}, \quad \delta_{\text{F}} \equiv \frac{\langle \mathbf{E}^2 + \mathbf{B}^2 \rangle}{6H^2 M_{\text{P}}^2}. \quad (3.27)$$

The first parameter, δ_{KG} quantifies the backreaction (BR) in the Klein–Gordon equation, Eq. (3.11) and indicates the importance of gauge-field friction over the usual Hubble friction during inflation. The second parameter, δ_{F} , instead indicates the relative importance of the electromagnetic energy density, ρ_{EM} , to the total energy density $\rho_{\text{tot}} = 3M_{\text{P}}^2 H^2$. As ρ_{EM} evolves as radiation, we require $\delta_{\text{F}} \lesssim 1/2$ to ensure inflation, which is to say, the potential energy of the inflaton needs to dominate (c.f. Eq. (3.14)). The weak BR regime is then classified through $\delta_{\text{KG}} \ll 1$ and $\delta_{\text{F}} \ll 1$, while the strong-BR regime corresponds to $\delta_{\text{KG}} \gtrsim 1$ with $\delta_{\text{F}} \lesssim 1/2$. In what follows, we now discuss these two regimes in turn, beginning with the simpler of the two: the weak-backreaction regime.

3.3.1 The regime of weak backreaction

Neglecting gauge-field backreaction, the inflaton’s EOM in Eq. (3.11) simply reduces to the conventional Klein–Gordon equation for a homogeneous scalar field as in Eq. (1.22). Therefore, the evolution of the inflaton field can simply be described by the standard slow-roll attractor discussed in Sec. 1.2. This then allows for a fully analytical description of axion inflation in the weak-backreaction regime.

Assuming the inflaton evolves on the slow-roll attractor, we have that the inflationary spacetime is approximately given by de Sitter space with a constant Hubble rate, H , while the inflaton velocity, $\dot{\phi}$, is also nearly constant (cf. Sec. 1.2). This then implies that $\eta \simeq -1/(aH)$, while the instability parameter ξ is also nearly constant. Under these assumptions, the mode equation for a gauge-field mode $A_{\lambda}(\eta, k)$ can be solved exactly:

$$A_{\lambda}(\eta, k) = \frac{1}{\sqrt{2k}} e^{\frac{\lambda\pi\xi}{2}} W_{-i\lambda\xi, \frac{1}{2}}(2ik\eta). \quad \text{for } H = \text{const.} \ \& \ \xi = \text{const.} \quad (3.28)$$

Here, $W_{\kappa, \mu}$ is the Whittaker- W function (see App. A). From this expression, we can also explicitly recognize the role played by the instability parameter, ξ , which controls the exponential enhancement of one gauge-field polarization over the other.

In the weak-backreaction regime, we then have that gauge-field production is entirely determined through ξ . This also carries over to the expectation values for the amplified electromagnetic background in Eq. (3.25), which, are also simply functions of ξ :

$$\langle \mathbf{E} \cdot \mathbf{B} \rangle = H^4 \Upsilon_0(\xi), \quad \langle \mathbf{E}^2 + \mathbf{B}^2 \rangle = H^4 \bar{\rho}_{\text{EM},0}(\xi). \quad (3.29)$$

To find explicit expressions for these functions, one can rely on a commonly adopted approximation of the mode function in Eq. (3.28) for the tachyonically enhanced polarization mode [54–56]

$$A_{\text{sign}(\xi)}(\eta, k) \simeq \frac{1}{\sqrt{2k}} \left(\frac{k}{2|\xi|aH} \right) e^{\pi|\xi|-2\sqrt{2|\xi|k/(aH)}}. \quad (3.30)$$

which holds for $|\xi| \gg 1$ and $(8|\xi|)^{-1} \ll k/aH \ll 2|\xi|$. Evaluating the integrals in Eq. (3.25) using this approximation, while neglecting the exponentially suppressed gauge-field mode for $\lambda = -\text{sign}(\xi)$, yields [1]

$$\Upsilon_0(\xi) = -\text{sign}(\xi) \frac{7!}{2^{21}\pi^2} \frac{e^{2\pi|\xi|}}{|\xi|^4}, \quad \bar{\rho}_{\text{EM},0}(\xi) = \frac{7!}{2^{21}\pi^2} (1 + 1.1|\xi|^2) \frac{e^{2\pi|\xi|}}{|\xi|^5} \quad (3.31)$$

From the exponential dependence of these expressions on $|\xi|$, we can recognize explicitly how potentially abundant gauge-field production during axion inflation can be for large $|\xi|$. This also justifies our earlier idea that ξ is a gauge-field-production parameter.

The approximation of constant ξ and H we have made so far is of course an idealization, which is not realized on a realistic slow-roll trajectory. Generically, $\dot{\varphi}$ slowly increases through the course of inflation, while H slowly decreases. Consequently, also $|\xi|$ grows during inflation, and, therefore, so does gauge-field production. This continuous amplification of the electromagnetic background can result in a multitude of interesting phenomenological implications (see Sec. 3.4). However, this growth also entails that the dynamical impact of the gauge fields consistently increases as the system evolves.

A simple way to see this increased dynamical effect explicitly is to evaluate the BR parameters in Eq. (3.27) using Eqs. (3.29) and (3.31) [4]:

$$\delta_{\text{KG}} \simeq 4.1 \times 10^{-5} \left(\frac{\beta H}{M_{\text{P}}} \right)^2 \frac{e^{2\pi|\xi|}}{|\xi|^5}, \quad \delta_{\text{F}} \simeq 4.1 \times 10^{-5} \left(\frac{H}{M_{\text{P}}} \right)^2 \frac{e^{2\pi|\xi|}}{|\xi|^5} (1 + 1.1|\xi|^2). \quad (3.32)$$

The notable exponential sensitivity of these parameters on $|\xi|$ highlights that even a small change in the instability parameter on the slow-roll attractor can entail that a system which starts out in weak BR may transition to a regime of strong BR, where gauge-field friction effects are no longer negligible. In this sense, backreaction effects are for some configurations rather an inevitability of axion inflation than a fringe case, and we need to consider how the inflationary system responds to this amplified background of electromagnetic radiation.

3.3.2 The regime of strong backreaction

At first, it may not be evident why the onset of BR effects leads to a new inflationary regime of strong BR, rather than simply a breakdown of inflation itself. After all, not only the friction term $\langle \mathbf{E} \cdot \mathbf{B} \rangle$ is sourced by the rolling inflaton field but also the electromagnetic energy density, ρ_{EM} , whose growth can threaten the domination of the inflaton's potential energy density. However, friction turns out to be the more important dynamical effect in the lead up from weak BR, as we can see by studying Eq. (3.32). These imply that $\delta_{\text{F}} \simeq \beta^2 \delta_{\text{KG}}$. Therefore, a dynamical system in the regime of weak BR can indeed transition to an inflationary stage with strong BR, $\delta_{\text{KG}} \sim 1$ and $\delta_{\text{F}} \lesssim 1$, as long as the axion–vector coupling β is large.

Once electromagnetic dissipation through $\langle \mathbf{E} \cdot \mathbf{B} \rangle$ becomes important, the standard slow-roll attractor we relied on for weak BR breaks down, and the assumption that ξ changes adiabatically slowly ceases to be true. We can then no longer rely on the analytical expressions for the mode functions in Eq. (3.28), or assume a direct dependence of $\langle \mathbf{E} \cdot \mathbf{B} \rangle$ and ρ_{EM} on ξ as in Eq. (3.29). Originally, ANBER AND SORBO [55] suggested the existence of a new attractor in the regime of strong BR, where gauge-field friction is stable and can support a constant value of ξ . However, this supposed attractor is not reached in numerical simulations of the dynamical transition from weak to strong BR [106–109], and, in fact, appears to be unstable [131]. Chapter 5 of this thesis is entirely dedicated to the discussion of this *Anber–Sorbo solution* and its possible instability.

At the heart of the unstable dynamics of strong BR is the emergence of a second timescale in the system, a retarded response between changes in ξ and gauge-field production [109]. A hint at this retardation can in fact be seen in the spectrum shown in Fig. 3.1. The important feature therein is that the peak of the amplified mode $|A_-(t, k)|$ is not at k_{thr} , but at a mildly lower wavenumber, which crossed the threshold scale at an earlier point in time, $k_{\text{thr}}(t - \Delta t)$. This implies that backreaction terms, like $\langle \mathbf{E} \cdot \mathbf{B} \rangle$, carry a memory from the state of the physical system at $t - \Delta t$. This retardation can prevent the dynamical system from stabilizing gauge-field friction and can trigger highly nonlinear behavior.

While the emergence of this nonlinear evolution has been understood analytically [109, 131], its full dynamical effects have so far only been successfully resolved using numerical techniques. There are three main numerical methods for this purpose: (i) *the mode-by-mode approach* [83, 85, 87, 88, 106–109, 375], in which one directly solves the coupled set of equations (3.26) by evolving individual gauge-field modes $A_\lambda(t, k)$ in time; (ii) *lattice simulations* [376–381] which solve the equations of motion for axion inflation directly on a discretized lattice; and (iii) the *gradient-expansion formalism* (or GEF) [128, 133, 372], which forgoes a description of gauge-field production in momentum space by directly establishing a tower of differential equations governing the evolution of the backreaction terms, $\langle \mathbf{E} \cdot \mathbf{B} \rangle$, $\langle \mathbf{E}^2 \rangle$ and $\langle \mathbf{B}^2 \rangle$. The GEF is the central method we rely on in this thesis, and chapter 4 is fully dedicated to its discussion.

All these methods reveal a similar dynamical structure which is characterized by chaotic oscillations in ξ . Qualitatively, these oscillations arise as follows. At first, strong friction through $\langle \mathbf{E} \cdot \mathbf{B} \rangle$ brakes the inflaton, and thus ξ decreases. This entails a loss in gauge-field production, and therefore, gauge-field friction diminishes. However, the reduction in electromagnetic dissipation allows the inflaton field to accelerate again, ξ grows, friction increases, and the cycle repeats. Over multiple oscillations, the additional friction from the gauge field however accumulates, and results in a loss of kinetic energy in the inflaton field. This keeps a system in strong BR from ever reaching the usual end of inflation with $\dot{\varphi}^2 \simeq V(\varphi)$. Instead, inflation is prolonged, and ultimately ends once gauge-field production catches up with the inflaton’s potential energy density, $\rho_{\text{EM}} \simeq V(\varphi)$.

Physically, these effects of backreaction can also be understood as gauge-field friction inducing a stuttering motion of the inflaton field while it is descending its potential. In fact, it may happen that the inflaton’s velocity briefly changes sign in a regime of strong BR, corresponding to the inflaton briefly rolling backwards and ascending its potential.

To illustrate these dynamics, and how they compare to those of weak backreaction, we show two different evolution histories for the energy densities and backreaction parameters in Fig. 3.3. In the left panel, backreaction effects never become important (δ_{KG} remains below unity), and the system reaches the end of inflation on the slow-roll trajectory, $\dot{\varphi}^2 \simeq V(\varphi)$. Meanwhile, the system on the right starts out in a regime of weak BR, but friction effects become important before inflation can end (δ_{KG} crosses unity). This system then undergoes a period of strong BR, which prolongs inflation for several e -folds past the expected end of slow-roll inflation (vertical dashed line) and only ends for $\rho_{\text{EM}} \simeq V(\varphi)$.

Seeing as the dynamics of strong BR are highly nonlinear, such a qualitative description is the extent to which strong-backreaction effects can be characterized. To round off its discussion, we should also note that recent developments in lattice simulations have revealed a new feature that can emerge in the regime of strong BR (see Refs. [377–380]). These lattice studies demonstrate that gauge-field production can efficiently source inhomogeneities in the inflation field, $\delta\phi(t, \mathbf{x})$, which we have assumed to be dynamically unimportant in our discussion thus far (cf. Sec. 3.2). In the strong BR regime, these inhomogeneities may however backreact on gauge-field production itself. This can lead to a regime of *inhomogeneous backreaction* whose main characteristic is a damping of the oscillatory behavior of ξ we just described. However, qualitatively, this inhomogeneous backreaction still leads to a prolonging of inflation, and an ultimate domination of

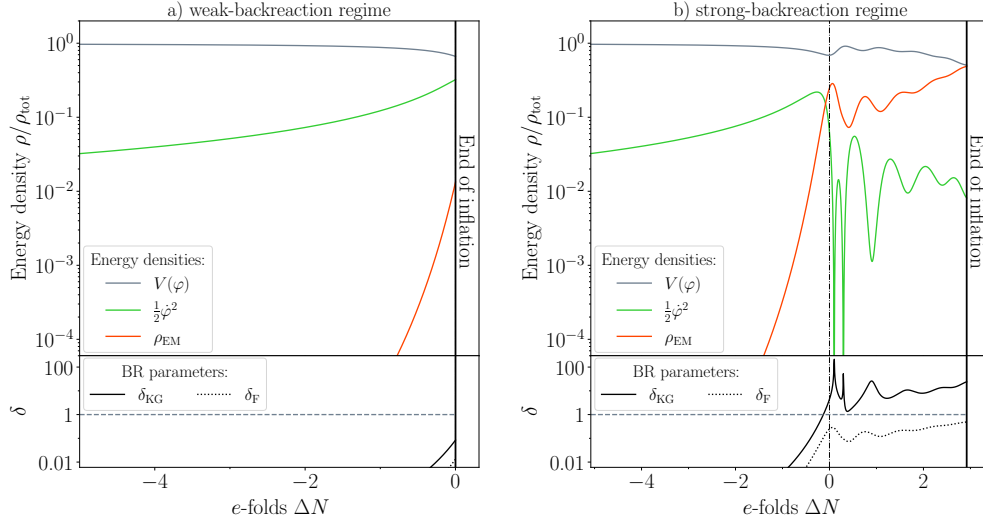


FIGURE 3.3: Two different evolution histories for axion inflation. The top panels show the change in the inflaton’s potential (gray) and kinetic (green) energy density alongside the electromagnetic energy density (red) normalized to $\rho_{\text{tot}} = 3H^2 M_{\text{P}}^2$. The lower panel shows the evolution of the backreaction parameters δ_{KG} (solid black) and δ_{F} (dotted black) in Eq. (3.27). On the x-axis, we show the number of e -folds, ΔN , until the expected end of inflation on the usual slow-roll trajectory. **Panel (a):** The system remains in the weak-backreaction regime until the end of inflation at $\Delta N = 0$. **Panel (b):** The system enters the strong-backreaction regime before $\Delta N = 0$ and inflation is prolonged by three e -folds. These results are obtained using the GEF for a quadratic potential $V(\phi) = m^2 \phi^2/2$ with a mass $m = 6.16 \times 10^{-6} M_{\text{P}}$ and for two different coupling constants, $\beta = 13.5$ (left), and $\beta = 15$ (right). Both systems are initialized in the weak-backreaction regime at $\Delta N = -60$ (cf. Sec. 4.2.2).

the electromagnetic energy density. In the next chapter, we shall discuss the regime of inhomogeneous backreaction in more detail, after we have established the GEF as method for studying strong-backreaction dynamics in axion inflation.

In summary, we may conclude that axion inflation features a rich dynamical structure which is the topic of ongoing research. In particular, these dynamics require numerical methods to resolve them, the development and refinement of which is the purpose of the next part of this thesis.

3.4 Implications of gauge-field production

Having discussed the dynamical production of gauge fields in axion inflation, we now close this chapter by briefly elaborating on their possible phenomenological implications. This is mainly intended as an overview and to establish fundamental concepts that we can draw on for later discussions. Meanwhile, many of these topics will be discussed in more detail in dedicated chapters, where we present our own work on these topics.

3.4.1 Sourced primordial perturbations

The presence of an amplified background of electromagnetic fields during inflation can directly source correlated scalar and tensor perturbations [55, 56, 58, 382]. These sourced perturbations then add a novel contribution to the total scalar and tensor power spectrum during inflation in addition to the usual contribution of vacuum metric and inflaton perturbations we have already discussed in chapter 1.

For gauge-field production in the weak-backreaction regime, there are established analytical formulas for the total scalar and tensor power spectrum which parametrize the gauge-field-induced contribution in terms of the gauge-field production parameter ξ . For the scalar power spectrum, one finds [56, 57]

$$\mathcal{P}_\zeta(k, \eta) \simeq \left(\frac{H^2}{2\pi\dot{\phi}} \right)^2 \left[1 + 7.5 \times 10^{-5} \left(\frac{H^2}{2\pi\dot{\phi}} \right)^2 \frac{e^{4\pi|\xi|}}{|\xi|^6} \right], \quad \text{for } -k\eta \ll 1 \text{ \& } \xi \gg 1 \quad (3.33)$$

while the tensor power spectrum is given by [58]

$$\mathcal{P}_h(k, \eta) \simeq \left(\frac{H}{\pi M_{\text{P}}} \right)^2 \left[2 + 8.6 \times 10^{-7} \left(\frac{H}{M_{\text{P}}} \right)^2 \frac{e^{4\pi|\xi|}}{|\xi|^6} \right], \quad \text{for } -k\eta \ll 1 \text{ \& } \xi \gg 1 \quad (3.34)$$

with the first term in each equation being the standard vacuum contributions in Eqs. (1.77) and (1.82), while the second term is due to gauge fields. As previously discussed in Sec. 3.3.1, ξ generally grows in the regime of weak backreaction. Following the logic that perturbations with smaller wavelengths are sourced at later times during inflation, therefore means that the gauge-field-induced contribution is more relevant on smaller scales than on larger scales.

The scalar perturbations sourced by the electromagnetic background can be highly non-Gaussian [56, 57, 339, 383]. Meanwhile, the degree of non-Gaussianity at CMB scales is strongly constrained [15, 49], which therefore limits the efficiency of gauge-field production at these scales. The corresponding bound is conventionally expressed in terms of the instability parameter as $|\xi_{\text{CMB}}| \lesssim 2.5$ [56, 57]. Meanwhile, ξ may dynamically grow beyond this limit at late times during inflation. As a result, the scalar power spectrum can be enhanced at small scales, which can then indirectly lead to the formation of primordial black holes [89, 90] or sourced SIGWs [109] as discussed in Sec. 1.5.

Because of the limits on scalar non-Gaussianity, gauge-field-induced tensor perturbations at CMB scales are likely unimportant, thus they do not affect the conventional results for the tensor-to-scalar ratio r (one exception to this are k-fation scenarios of axion inflation, see Ref. [384]). However, because the sourced tensor perturbations are also enhanced towards smaller scales, they may source a stochastic background of gravitational waves (SGWB) that could be observable with current or next-generation gravitational wave (GW) experiments [51, 82–88]. In particular, the helical nature of gauge-field production during axion inflation also transfers to these sourced GWs, which are thus polarized [58]. The helical nature of this SGWB could be used to distinguish this signal from other potential sources [51, 58]. The efficiency of GW production via gauge fields is however limited, as the GWs would act as an additional species of dark radiation beyond the measured levels of SM radiation at the time of BBN or CMB emission [380, 385, 386].

In part III of this thesis, we study both the sourcing of scalar and tensor perturbations during axion inflation, which is where we also discuss how to compute the sourced contributions to their power spectra in the regime of strong gauge-field backreaction, where the analytical formulas in Eqs. (3.33) and (3.34) no longer apply.

3.4.2 Schwinger pair production

An important effect which our discussion of gauge-field production and backreaction during axion inflation in the previous sections has not taken into consideration is the possibility that the amplification of electromagnetic fields via the inflaton field could itself be a channel for the production of charged fermions or scalars via Schwinger pair-creation [120, 121]. This production channel can be relevant if any scalars or fermions are charged under the Abelian gauge group of

the gauge field A_μ appearing in Eq. (3.7), which is for instance the case if one considers that this Abelian gauge field corresponds to the SM hypercharge field [128, 130].

In cases where the Schwinger effect is efficient at producing charged particles, it can lead to significant modifications to the dynamics of axion inflation: the pair-creation process can effectively drain energy from the gauge-field background [127], which inhibits the potential for gauge-field backreaction on the inflaton dynamics [104, 128, 129, 134]. This can be understood by considering that the Schwinger effect induces an effective charge current $\mathbf{J}$, which enters into the existing Maxwell's equations for axion inflation as a modification to Ampère's law (see e.g., [133])

$$\dot{\mathbf{E}} + 2H\mathbf{E} - \frac{1}{a}\text{rot}\mathbf{B} + \frac{\beta}{M_{\text{P}}}\dot{\phi}\mathbf{B} + \mathbf{J} = 0. \quad (3.35)$$

In addition to its dynamical implications, the Schwinger effect as a channel for creating charged particles, is also phenomenological relevant. For instance, the combined production of helical gauge fields and fermions in axion inflation has been considered as a source of CP violation in the context of baryogenesis scenarios [102–105]. Other authors have considered Schwinger pair production as a mechanism for inflationary dark-matter production [125, 126]. Furthermore, charged fields, like fermions, can also induce GWs [122, 123], which then would add another sourced contribution to the tensor power spectrum from axion inflation. However, this contribution, to our knowledge, has not yet been computed in the literature.

In chapter 6, we discuss how Schwinger pair creation during axion inflation can be modelled and how the GEF can be used to describe its dynamical implications. Meanwhile, in chapter 9, we consider how the predictions for gauge-field-induced GWs in axion inflation are modified once Schwinger pair creation has become effective at inhibiting the production of gauge fields.

3.4.3 Other effects

Before proceeding to discuss the dynamical implications of gauge-field production in more detail in the next part of this thesis, let us also give a brief overview of other effects of gauge-field production which we shall not further discuss in this thesis.

For one, gauge-field production in axion inflation can be an efficient channel for preheating, as has been analyzed in several lattice studies [78–81]. Gauge fields that are created during the reheating era can then also further generate GWs, see Refs. [80, 385, 386].

The creation of helical magnetic fields has also been studied as a mechanism for generating primordial magnetic fields [53, 54, 106, 387]. Recent lattice simulations of axion inflation [380] indeed find that axion inflation may be a viable magnetogenesis scenario, as it could source magnetic fields with a sufficient amplitude to explain the evidence for large scale magnetic fields in our present Universe suggested by blazar observations [95–97]. However, the amplitude of these magnetic fields may be significantly reduced when accounting for the damping effects of charge carriers produced via the Schwinger effect [134].

Some authors have also considered that the gauge field can thermalize in models of axion inflation [388, 389]. However, existing lattice simulations have not observed this, at least in models of axion inflation in which the gauge field does not couple to charged fields. Meanwhile, these thermalization effects could be more important when gauge fields do couple to charged fields, for instance, to SM particles [388, 389].

Part II

The dynamics of axion inflation

Chapter 4

The gradient-expansion formalism

In this second part to this thesis, we introduce and develop the main tools which we use to describe the dynamical production of gauge fields during axion inflation. As we have seen in the last chapter, the model of axion inflation features a rich dynamical structure, where gauge fields are both efficiently sourced by the inflaton field, but can also directly backreact on its dynamics. This dynamical interplay between gauge-field production and backreaction needs to be consistently taken into account to determine the inflationary evolution during axion inflation, describe the efficiency of gauge-field production in this model, and to ultimately study their phenomenological implications. In this chapter, we introduce our main numerical tool for this purpose: the gradient-expansion formalism.

At its core, the gradient-expansion formalism (GEF) is a reformulation of the equations of motion of axion inflation: instead of describing gauge-field production and backreaction in terms of gauge-field mode functions, as we have done it in the previous chapter, the GEF entails the construction of a coupled set of ordinary differential equations (ODEs) that directly govern the evolution of the gauge-field expectation values responsible for backreaction effects. The original GEF formulation of axion inflation was developed in Ref. [128] and has been further developed and applied in Refs. [133, 372]. Meanwhile, the GEF has also been used in the study of other inflationary models featuring gauge-field production and backreaction effects [135, 390–392], for instance in models where the inflaton couples to gauge fields via a modified kinetic term as in the original Ratra model [393]. The GEF can also be adapted to describe the dynamical Schwinger pair creation of charge carriers during axion inflation [128, 130, 133], which we further discuss in chapter 6.

The main purpose of our subsequent discussion is to establish the fundamental idea behind the GEF, to illustrate how it can be used as a numerical tool to determine the dynamics of axion inflation, and to highlight its advantages and disadvantages with respect to other techniques. To this end, we focus on the simplest version of the GEF, which is adapted to the basic axion inflation model that we introduced in the last chapter. In the next two chapters, we will then expand upon and adapt this basic GEF version to more specific scenarios, but the basic ideas we present in this chapter will remain the backbone of these modified GEF versions. Also, in part III, the GEF we present here is needed to provide the necessary input for our discussion of gauge-field-sourced inflationary perturbations for axion inflation in the strong-backreaction regime.

Our presentation of the GEF in this chapter is structured as follows. In Sec. 4.1, we present the fundamental equations of motion that the GEF relies on to describe the dynamics of axion inflation. Following this, Sec. 4.2 presents the numerical strategy that we employ to solve these GEF equations. Here, we also discuss the standard dynamical scenario to which we apply the GEF in later parts of this thesis. In Sec. 4.3, we then reflect on the GEF's limitations and advantages compared to other numerical methods for studying axion inflation, in particular with a focus on lattice simulations. Finally, we end this chapter in Sec. 4.4 by giving an outlook regarding the applications of the GEF that we will consider in the subsequent parts of this thesis.

4.1 The GEF formulation of axion inflation

We start our discussion by presenting the basic GEF formulation of axion inflation. This means, we describe how the equations of motions for a homogeneous inflaton field $\varphi \equiv \langle \phi \rangle$, Eq. (3.11), the electric and magnetic field operators $\mathbf{E}$ and $\mathbf{B}$, Eq. (3.12), and the Friedmann equation (3.13), may be transformed into a finite set of ODEs for gauge-field expectation values like $\langle \mathbf{E}^2 \rangle$, $\langle \mathbf{B}^2 \rangle$ and $\langle \mathbf{E} \cdot \mathbf{B} \rangle$. In this, we follow the discussion presented in Ref. [128].

4.1.1 Reformulating the equations of motion of axion inflation

The starting point for the GEF technique is the definition of the dynamical quantities which are at the heart of its construction. They are a set of bilinear gauge-field expectation values of the electromagnetic field operators $\mathbf{E}$ and $\mathbf{B}$ that we shall refer to as the *GEF variables*:¹

$$\mathcal{E}^{(n)} = \frac{1}{a^n} \langle \mathbf{E} \cdot \text{rot}^n \mathbf{E} \rangle, \quad (4.1a)$$

$$\mathcal{G}^{(n)} = -\frac{1}{2a^n} \langle \mathbf{E} \cdot \text{rot}^n \mathbf{B} + \mathbf{B} \cdot \text{rot}^n \mathbf{E} \rangle, \quad (4.1b)$$

$$\mathcal{B}^{(n)} = \frac{1}{a^n} \langle \mathbf{B} \cdot \text{rot}^n \mathbf{B} \rangle, \quad (4.1c)$$

For $n = 0$, these GEF variables are simply the backreaction terms $\langle \mathbf{E}^2 \rangle$, $\langle \mathbf{E} \cdot \mathbf{B} \rangle$ and $\langle \mathbf{B}^2 \rangle$ which appeared ubiquitously throughout our discussion of axion inflation in the previous chapter. Thus, using these new variables, the Klein–Gordon equation (3.11) and the Friedmann equation (3.13) for axion inflation are simply,

$$\ddot{\varphi} + 3H\dot{\varphi} + V_{,\phi}(\varphi) = -\frac{\beta}{M_{\text{P}}} \mathcal{G}^{(0)}, \quad (4.2)$$

$$H^2 = \frac{1}{3M_{\text{P}}^2} \left[\frac{1}{2} \dot{\varphi}^2 + V(\varphi) + \frac{1}{2} (\mathcal{E}^{(0)} + \mathcal{B}^{(0)}) \right]. \quad (4.3)$$

Here, as before, H is the Hubble rate, V is the inflaton potential, and $V_{,\phi}$ denotes its gradient. The reason for introducing the whole generalized set of bilinear gauge-field expectation values in Eq. (4.1) is that they allow one to elegantly reformulate the entire dynamics encoded in Maxwell's equations (3.12) as a coupled set of differential equations for $\mathcal{E}^{(n)}$, $\mathcal{G}^{(n)}$, and $\mathcal{B}^{(n)}$:

$$\dot{\mathcal{E}}^{(n)} + (n+4)H\mathcal{E}^{(n)} - \frac{2\beta}{M_{\text{P}}} \dot{\varphi} \mathcal{G}^{(n)} + 2\mathcal{G}^{(n+1)} = S_{\mathcal{E}}^{(n)}, \quad (4.4a)$$

$$\dot{\mathcal{G}}^{(n)} + (n+4)H\mathcal{G}^{(n)} - \frac{\beta}{M_{\text{P}}} \dot{\varphi} \mathcal{B}^{(n)} - \mathcal{E}^{(n+1)} + \mathcal{B}^{(n+1)} = S_{\mathcal{G}}^{(n)}, \quad (4.4b)$$

$$\dot{\mathcal{B}}^{(n)} + (n+4)H\mathcal{B}^{(n)} - 2\mathcal{G}^{(n+1)} = S_{\mathcal{B}}^{(n)}. \quad (4.4c)$$

This is at the heart of the GEF: we exchange our original description of axion inflation in terms of mode functions and integrals (cf. chapter 3) for a new description in terms of coupled ODEs.

In its current form, Eq. (4.4) evidently require further explanation. For one, we need to explore the origin of the source terms (or *boundary terms*) $S_{\mathcal{X}}^{(n)}$ for $\mathcal{X} = \mathcal{E}, \mathcal{G}, \mathcal{B}$ appearing on its right-hand side. Furthermore, each differential equation describing the evolution of $\mathcal{X}^{(n)}$ for a fixed n depends

¹The symmetrization in the definition of $\mathcal{G}^{(n)}$ is included since the operators $\text{rot}^n \mathbf{E}$ and $\text{rot}^m \mathbf{B}$ do not commute for $n, m > 1$.

on GEF variables of order $n + 1$.² So, our current GEF formulation of axion inflation is one of infinitely many coupled differential equations, making them unsuitable for practical applications. In what follows, we will therefore first explain the origin and interpretation of the boundary terms, and then describe a scheme to analytically close the infinite tower of GEF equations in Eq. (4.4) at some finite order n . In this way, we obtain a finite set of ODEs which we can use in practice to study both the dynamical production of gauge fields during axion inflation, and their backreaction on the inflaton dynamics.

4.1.2 Boundary terms

The origin of the boundary terms $S_{\mathcal{X}}^{(n)}$ is most clearly revealed by momentarily returning to a momentum-space description of the GEF variables $\mathcal{X}^{(n)}$. At their core, the GEF variables defined in Eq. (4.1) are simply bilinear expectation values of the gauge-field operator $\mathbf{A}$ (cf. Eq. (3.15)). Thus, we may also directly represent them as integrals over the gauge-field mode functions, $A_\lambda(t, k)$:

$$\mathcal{E}^{(n)} = \int_0^{k_h(t)} \frac{dk}{k} \frac{k^{n+3}}{2\pi^2 a^{n+2}} \sum_\lambda \lambda^n |\dot{A}_\lambda(t, k)|^2, \quad (4.5a)$$

$$\mathcal{G}^{(n)} = \int_0^{k_h(t)} \frac{dk}{k} \frac{k^{n+4}}{2\pi^2 a^{n+3}} \sum_\lambda \lambda^{n+1} \text{Re}[\dot{A}_\lambda(t, k) A_\lambda^*(t, k)], \quad (4.5b)$$

$$\mathcal{B}^{(n)} = \int_0^{k_h(t)} \frac{dk}{k} \frac{k^{n+5}}{2\pi^2 a^{n+4}} \sum_\lambda \lambda^n |A_\lambda(t, k)|^2. \quad (4.5c)$$

These equations follow directly from our definition of the expectation value $\langle \cdot \rangle$ for the quantized gauge field $\mathbf{A}$, as we have described it in Sec. 3.2.2. In particular, this means that we regulate these UV-divergent integrals by introducing the instability scale k_h in Eq. (3.24) as a UV cutoff. The idea behind this regularization procedure is as before; k_h qualitatively separates quantum vacuum modes from modes that have been excited by the rolling inflaton field, and which can therefore contribute to a background configuration of classical gauge fields.

The time dependence of the UV regulator $k_h(t)$ is the reason the boundary terms appear in Eq. (4.4). To construct these equations for the GEF variables $\mathcal{X}^{(n)}$, we need to take the time derivative of the expectation values for a quantum operator, for instance $\langle \mathbf{E} \text{rot}^n \mathbf{B} \rangle$. However, Maxwell's equations (3.12) merely describes the time variation of the operators $\mathbf{E}$ and $\mathbf{B}$. The boundary terms are instead needed to capture the time dependence of $\langle \cdot \rangle$ itself, which is due to its dependence on the UV regulator, k_h . Physically, we can understand the boundary terms as accounting for the constant influx of vacuum modes $A_\lambda(t, k)$ which are continuously excited out of their vacuum configuration by the rolling of the inflaton field during inflation.

Let us elaborate on this point to derive the mathematical structure of the boundary terms $S_{\mathcal{X}}^{(n)}$. Following Eq. (4.5), any GEF variable $\mathcal{X}^{(n)}$ can be represented as an integral of the form

$$\mathcal{X}^{(n)}(t) = \int_0^{k_h(t)} \frac{dk}{k} \frac{d\mathcal{X}^{(n)}(k, t)}{d \log k}, \quad (4.6)$$

²For notational simplicity, we use $\mathcal{X}$ as a collective label for discussions which equally apply to all three bilinear quantities in Eq. (4.1). For example, $\mathcal{X}^{(n)}$ may represent either $\mathcal{E}^{(n)}$, $\mathcal{G}^{(n)}$ or $\mathcal{B}^{(n)}$.

where $d\mathcal{X}^{(n)}(k, t)/d \log k$ denotes the spectrum of $\mathcal{X}^{(n)}$. The time derivative of $\mathcal{X}^{(n)}$ then takes two contributions: one from the time dependence of the spectrum, and one from the time dependence of k_h :

$$\dot{\mathcal{X}}^{(n)}(t) = \int_0^{k_h(t)} \frac{dk}{k} \frac{d}{dt} \left(\frac{d\mathcal{X}^{(n)}(k, t)}{d \log k} \right) + \underbrace{\frac{d \log k_h(t)}{dt} \frac{d\mathcal{X}^{(n)}(k, t)}{d \log k} \Big|_{k=k_h(t)}}_{\equiv S_{\mathcal{X}}^{(n)}}. \quad (4.7)$$

Maxwell's equations allows us to express the first term through other GEF variables, such that we can construct the left-hand side of the GEF equation (4.4), while the second contribution constitutes the boundary term $S_{\mathcal{X}}^{(n)}$ on the right-hand side of Eq. (4.4).

Formally, the boundary terms for the three types of GEF variables, $\mathcal{E}^{(n)}$, $\mathcal{G}^{(n)}$ and $\mathcal{B}^{(n)}$, are given by

$$S_{\mathcal{E}}^{(n)}(t) = \frac{d \ln k_h(t)}{dt} \frac{k_h(t)}{2\pi^2} \left(\frac{k_h(t)}{a(t)} \right)^{n+2} \sum_{\lambda=\pm 1} \lambda^n |\dot{A}_{\lambda}(t, k_h(t))|^2, \quad (4.8a)$$

$$S_{\mathcal{G}}^{(n)}(t) = \frac{d \ln k_h(t)}{dt} \frac{k_h(t)}{2\pi^2} \left(\frac{k_h(t)}{a(t)} \right)^{n+3} \sum_{\lambda=\pm 1} \lambda^{n+1} \text{Re} [A_{\lambda}^*(t, k_h(t)) \dot{A}_{\lambda}(t, k_h(t))], \quad (4.8b)$$

$$S_{\mathcal{B}}^{(n)}(t) = \frac{d \ln k_h(t)}{dt} \frac{k_h(t)}{2\pi^2} \left(\frac{k_h(t)}{a(t)} \right)^{n+4} \sum_{\lambda=\pm 1} \lambda^n |A_{\lambda}(t, k_h(t))|^2. \quad (4.8c)$$

Evidently, these expressions explicitly depend on the mode function $A_{\lambda}(t, k)$. In particular, they are the mode functions with wavenumber $k = k_h(t)$, i.e., those modes which are crossing the instability scale at time t .

At first glance, the fact that these boundary terms appear in Eq. (4.4) entails that the GEF formulation of axion inflation can not entirely rid itself of mode functions. We can however analytically approximate the mode functions $A_{\lambda}(t, k_h)$ to circumvent this mixing between the GEF and the mode-function formulation of axion inflation. The important observation is that the mode functions appearing in Eq. (4.8) are, per definition of k_h , necessarily still in a near vacuum-like state. In particular, the fact that $k_h(t)$ is constructed as a monotonically increasing function (cf. Eq. (3.24)) means that a mode $A_{\lambda}(t, k_h(t))$ has not been subjected to the tachyonically enhancing effects of the inflaton field at any prior times $t' < t$. This then suggests that this gauge mode is not yet sensitive to the full dynamical state of the inflationary system at time t . Therefore, we may suitably approximate $A_{\lambda}(t, k_h(t))$ by assuming that it has thus far changed adiabatically slowly out of the Bunch–Davies vacuum. This assumption motivates us to describe $A_{\lambda}(t, k_h(t))$ in terms of the analytical expression for the gauge-mode functions in Eq. (3.28), which are valid for an adiabatically slowly changing system.

Following these considerations, we can then approximate the boundary terms in Eq. (4.8) as functions of time which are independent of the exact mode functions $A_{\lambda}(t, k)$,

$$S_{\mathcal{X}}^{(n)}(t) = \frac{d \ln k_h(t)}{dt} \frac{1}{4\pi^2} \left(\frac{k_h(t)}{a(t)} \right)^{n+4} \sum_{\lambda=\pm 1} \lambda^n \mathcal{X}_{\lambda}(\xi(t)), \quad \mathcal{X} = \mathcal{E}, \mathcal{G}, \mathcal{B}. \quad (4.9)$$

Here, ξ is the instability parameter defined in Eq. (3.20) and the spectral functions $\mathcal{X}_\lambda$ are

$$\mathcal{E}_\lambda(\xi) = \frac{e^{\pi\lambda\xi}}{4|\xi|^2} \left| (2i|\xi| - i\lambda\xi)W_{-i\lambda\xi, \frac{1}{2}}(-4i|\xi|) + W_{1-i\lambda\xi, \frac{1}{2}}(-4i|\xi|) \right|^2, \quad (4.10a)$$

$$\mathcal{G}_\lambda(\xi) = \frac{e^{\pi\lambda\xi}}{2|\xi|} \operatorname{Re} \left[W_{i\lambda\xi, \frac{1}{2}}(4i|\xi|)W_{1-i\lambda\xi, \frac{1}{2}}(-4i|\xi|) \right], \quad (4.10b)$$

$$\mathcal{B}_\lambda(\xi) = e^{\pi\lambda\xi} \left| W_{-i\lambda\xi, \frac{1}{2}}(-4i|\xi|) \right|^2, \quad (4.10c)$$

with $W_{\kappa, \mu}$ the Whittaker- W function appearing in Eq. (3.28). Note that, we have used the standard recurrence relationships for Whittaker functions that are given in App. A to arrive at the expressions for $\mathcal{E}_\lambda(\xi)$ and $\mathcal{G}_\lambda(\xi)$.

4.1.3 Approximate closing condition

Next, let us motivate a suitable condition to analytically close the infinite tower of differential equations given by Eq. (4.4). To this end, it is again useful to work with the integral representation of the GEF variables $\mathcal{X}^{(n)}$ in Eq. (4.5). This equation suggests that any $\mathcal{X}^{(n)}$ can be written as

$$\mathcal{X}^{(n)}(t) = \int_0^{k_h(t)} dk \left(\frac{k}{a} \right)^{n+4} \sum_\lambda \lambda^n \mathcal{I}_\lambda^{\mathcal{X}}(t, k), \quad (4.11)$$

where $\mathcal{I}_\lambda^{\mathcal{X}}(t, k)$ encodes the dependence of $\mathcal{X}^{(n)}$ on the mode functions $A_\lambda(t, k)$, e.g., for $\mathcal{B}^{(n)}$ we have $\mathcal{I}_\lambda^{\mathcal{B}}(t, k) = |A_\lambda(t, k)|^2/(2\pi^2)$. Considering the integral in Eq. (4.11) for a large value of the integer n , we can see that the integrand takes a dominant contribution from the term $(k/a)^{n+4}$ near the UV regulator, an effect which evidently is more pronounced the larger we take n . This observation suggests the following approximation:³

$$\mathcal{X}^{(n+1)} \approx \left(\frac{k_h}{a} \right)^2 \mathcal{X}^{(n-1)}, \quad \text{for } n \gg 1. \quad (4.12)$$

A more refined version of this simple truncation procedure was suggested in Ref. [372], which is simply a generalization of the one in Eq. (4.12),

$$\mathcal{X}^{(n+1)} \approx \sum_{l=1}^L (-1)^{l-1} \binom{L}{l} \left[\frac{k_h}{a} \right]^{2l} \mathcal{X}^{(n+1-2l)}, \quad \text{for } n \gg 1. \quad (4.13)$$

For $L = 1$, this condition simply reproduces Eq. (4.12).

Using Eq. (4.13) for some finite integer n_{cut} , the differential equation for the GEF variables $\mathcal{X}^{(n_{\text{cut}})}$ in Eq. (4.4) then no longer depends on $\mathcal{X}^{(n_{\text{cut}}+1)}$, but can instead be entirely expressed through terms with $n < n_{\text{cut}}$. Thus, this conditions allows us to truncate the infinite tower of differential equations in Eq. (4.4) at n_{cut} . For practical applications of the GEF, a typical truncation order is $n_{\text{cut}} \sim 100$, while Ref. [372] suggests $L \sim 10$ for the series expansion in Eq. (4.13).

With this analytical closing condition established, we therefore have successfully transformed the full system of coupled integro-differential equations that describe the dynamics of axion

³The polarization structure in Eq. (4.11) enforces that $\mathcal{X}^{(n+1)}$ with even (odd) n can only be related to $\mathcal{X}^{(n-1)}$, not $\mathcal{X}^{(n)}$. This can also be understood by noting that any such approximate equality should relate terms that transform in the same way under a parity transformation [390].

inflation, Eqs. (3.11), (3.13), and (3.18), into a finite set of coupled ODEs, Eqs. (4.2) to (4.4). These ODEs are then in a suitable form to be solved numerically.

4.2 Solving the GEF equations

Having established the practical formulation of the GEF, we now continue with a technical exposition of the numerical strategy used to solve the system of equations (4.2) to (4.4). This reflects the approach implemented in the Python code developed alongside this thesis [7], and is intended to highlight the key ingredients required for a stable and convergent solution of the GEF equations. Having discussed this general strategy, we then consider its concrete application to a standard benchmark scenario which we will return to as a point of reference throughout this thesis.

4.2.1 The GEF algorithm

The numerical procedure for resolving the dynamics of axion inflation using the GEF proceeds in three conceptual stages. At first, the physical system needs to be specified. For axion inflation, one requires the functional shape of the inflaton potential, $V(\phi)$, the strength of the axion–vector coupling β , and initial data on the inflaton field, $\varphi(0)$, $\dot{\varphi}(0)$, and the GEF variables. Furthermore, the truncation scheme outlined in Eq. (4.13) requires an initial choice for n_{cut} and L .

With the physical setup defined, Eqs. (4.2) to (4.4) can be solved using standard adaptive Runge–Kutta methods (e.g., RK45). In this way, one determines the full evolution of the dynamical background including the Hubble rate, the inflaton field, and the backreaction parameters $\mathcal{E}^{(0)}$, $\mathcal{G}^{(0)}$ and $\mathcal{B}^{(0)}$. Once the dynamical evolution of the background is known, one can also *a posteriori* determine the evolution of the gauge-field spectrum in terms of the mode functions $A_\lambda(t, k)$. The mode equation for axion inflation, Eq. (3.18), can be solved for a discrete set of gauge-field momenta, k_i . Each mode $A_\lambda(t, k_i)$ is initialized in the Bunch–Davies vacuum, Eq. (3.21), at a sufficiently early initial time t_i such that $k_i \gg k_h(t_i)$ is ensured. More details on how to solve a mode equation of this type are given in App. B.

The success of this numerical procedure can be validated using an internal error estimation scheme, which builds on the dual determination of the gauge-field evolution, once in terms of the gauge-field expectation values, $\mathcal{E}^{(n)}$, $\mathcal{G}^{(n)}$ and $\mathcal{B}^{(n)}$, and then again in terms of the gauge-mode spectrum, $A_\lambda(t, k)$. To assess the convergence of the GEF, one can determine $\mathcal{X}^{(0)}$ by directly integrating the spectrum of mode-function $A_\lambda(t, k)$ using Eq. (4.5). Denoting this second determination of $\mathcal{X}^{(0)}$ in terms of mode functions as $\mathcal{X}_{\text{MbM}}^{(0)}$, and the original determination by the GEF as $\mathcal{X}_{\text{GEF}}^{(0)}$, the convergence of the GEF result can be assessed by computing their relative deviations:

$$\epsilon_{\mathcal{X}}(t) = \left| 1 - \frac{\mathcal{X}_{\text{GEF}}^{(0)}(t)}{\mathcal{X}_{\text{MbM}}^{(0)}(t)} \right|, \quad \text{for } \mathcal{X} = \mathcal{E}, \mathcal{G}, \mathcal{B}. \quad (4.14)$$

This error estimate allows for an internal validation of the consistency between the GEF solution and the corresponding mode spectrum.⁴

This validation scheme is particularly relevant because of the approximate closing procedure that we use to attain a finite system of ODEs. When the system dynamics become strongly nonlinear, as is the case in the strong-backreaction regime of axion inflation, the closing condition in Eq. (4.13) can break down. For instance, this can happen in the case of short-timescale oscillations in the

⁴In practice, we use two criteria to claim a consistent GEF run: (a) the error at the final time of the numerical simulation, $\epsilon_{\mathcal{X}}(t_{\text{end}})$, should remain below a threshold (e.g., 10%); (b) the RMS error over all times t_i for N data points, $\epsilon_{\mathcal{X}}^{\text{RMS}} = \sqrt{\sum_i (\epsilon_{\mathcal{X}}(t_i))^2 / N}$, should remain below this same threshold value.

GEF variables $\mathcal{X}^{(n)}$, which typically appear in certain regions of the parameter space for axion inflation featuring large couplings β , alongside a low inflationary scale, H . A breakdown of the condition in Eq. (4.13) can destabilize the highest-order GEF variables, $\mathcal{X}^{(n_{\text{cut}})}$, whose dynamical value begins to deviate from its true value. These deviations can, over time, propagate through the system of equations (4.4), and ultimately impact the evolution of the backreaction parameters, $\mathcal{E}^{(0)}$, $\mathcal{G}^{(0)}$ and $\mathcal{B}^{(0)}$. This leads to a breakdown of the entire dynamical system, typically evidenced by positive-definite quantities, e.g., $\mathcal{E}^{(0)}$ and $\mathcal{B}^{(0)}$, becoming negative.

For the case where such a breakdown occurs, we devised a self-correction algorithm for the GEF in Ref. [1]. The algorithm proceeds as follows: (i) compute the mode spectrum $A_\lambda(t, k)$ based on the GEF solution; (ii) use the error estimator in Eq. (4.14) to identify the last valid state of the system at some time t_r ; (iii) compute the terms $\mathcal{X}^{(n)}(t_r)$ to all orders in n using the mode spectrum $A_\lambda(t_r, k)$; (iv) reinitialize the dynamical system at t_r using the corrected values $\mathcal{X}^{(n)}(t_r)$. The successive use of this self-correction procedure ultimately allows one to consistently determine the full system's evolution. Moreover, an increase of the truncation n_{cut} may also be necessary.

Despite the need for this self-correction algorithm, a full computation of the background dynamics and the gauge-mode spectrum is numerically cheap given that we are merely solving a coupled set of ODEs.⁵

Our Python package, the *Gradient Expansion Formalism Factory* (GEFF) [7] is a fully self-consistent implementation of the numerical strategy we have outlined above. In particular, it is designed to automatically detect a breakdown of the GEF truncation scheme, and it automatically applies the self-correction algorithm or increases n_{cut} if needed (cf. Fig. 1 in the GEFF manual [394]). The package also comes with built-in implementations of the GEF for the model of axion inflation. In chapter 6, we present an alternative version of the GEF, which is also designed to handle Schwinger pair production during axion inflation. This model is also implemented in the GEFF package. Beyond this, the code is designed to be modular, and allows for a simple implementation of the GEF also for other inflationary models featuring abundant gauge-field production and strong-backreaction effects.

4.2.2 Slow-roll initial conditions: a standard benchmark for the GEF

To illustrate the results that one may obtain using the above strategy, we now apply it to determine the full dynamical evolution of axion inflation starting from slow-roll initial conditions for a particular benchmark model of axion inflation.

Slow-roll initial conditions entail that the inflaton field is on a slow-roll trajectory at early times during inflation, e.g., when CMB scales exit the inflationary horizon, with gauge-field friction being negligible. In terms of the dynamical regimes of axion inflation described in the previous chapter, this means our system begins in the weak-backreaction regime of axion inflation allowing gauge-field production to gradually grow over time. This can also result in the onset of strong-backreaction effects if the inflaton has sufficiently amplified the electromagnetic background.

Slow-roll initial conditions are a conventional choice when studying gauge-field production in axion inflation (see, e.g., Refs. [109, 128, 378]). The reason for this is twofold. For one, slow-roll initial conditions are phenomenologically favorable because the CMB constraints on scalar non-Gaussianity impose stringent constraints on early gauge-field production [15, 49, 56, 57].

⁵Roughly, such a computation takes of the order of minutes for the GEF solution, and again of the order of minutes for the mode spectrum, in total between 5 – 15 min. The precise numbers of course depend on the available computational resources and the particularities of the model like coupling strength β and the inflaton potential $V(\phi)$. This estimate is representative of our own implementation in Python using our publicly available code in Ref. [394]. For comparison, Ref. [83] quotes ~ 30 min to obtain comparable results using their *Mathematica* implementation of the mode-by-mode approach to axion inflation.

Meanwhile, slow-roll initial conditions are relatively simple to realize with numerical methods. This is true for the GEF, but also when using the mode-by-mode approach (see, e.g., Ref. [83]) or simulating axion inflation on a lattice (see, e.g., Ref. [379]): assuming that gauge-field production has a negligible impact on the inflaton's evolution prior to initialization entails that all gauge-field backreaction terms, e.g., $\langle E^2 \rangle$, $\langle B^2 \rangle$, $\langle E \cdot B \rangle$, may be set to zero, while all gauge-field mode functions may be initialized in their vacuum state.

These considerations, in particular the constraints suggested by CMB measurements, also make slow-roll initial conditions relevant for this thesis, especially when we use the GEF for phenomenological analyses in later chapters. The following benchmark result therefore serves a dual purpose: we both use it to illustrate a typical GEF solution, but we can also use these results as a point of comparison in later chapters.

For our standard benchmark model, we rely on a simple quadratic potential for the inflaton field,

$$V(\phi) = \frac{1}{2}m^2\phi^2 \quad \text{with} \quad m = 6.16 \times 10^{-6} M_{\text{P}}, \quad (4.15)$$

where we set the inflaton mass to match the amplitude of scalar perturbations measured in the CMB (see chapter 1). This is a conventional benchmark which is studied in a variety of numerical analyses of axion inflation in the literature (see, e.g., Refs. [109, 128, 372, 376, 380]). Although a quadratic potential is ruled out by CMB measurements [14], it is a realistic approximation of the inflaton's potential toward the end of inflation. When studying gauge-field production, we are primarily interested in their effect at late times during inflation, where they may lead to an imprint on small physical scales, for instance, by inducing gravitational waves that are observable with interferometer or PTA experiments (see chapters 3 and 9). For such studies, a quadratic potential is a convenient and simple first choice.

Let us discuss how to initialize the GEF on the slow-roll trajectory for this potential. To obtain a complete perspective on the relevant inflationary dynamics, from the time CMB scales leave the horizon until the end of inflation, we initialize the inflaton field such that the system realizes at least 60 e -folds of inflation. We estimate this based on the dynamics of single-field inflation, allowing for backreaction effects to potentially prolong the duration of inflation. For a quadratic potential, this is achieved by setting ⁶

$$\varphi(0) = 15.55 M_{\text{P}}^2, \quad \dot{\varphi}(0) = -\sqrt{\frac{2}{3}}mM_{\text{P}}. \quad (4.16)$$

These initial conditions follow directly from a slow-roll analysis of the quadratic potential as in chapter 1 (see also, e.g., Ref. [138]). In line with the assumption that gauge-field production is initially absent, all GEF variables are initialized at zero:

$$\mathcal{E}^{(n)}(0) = 0, \quad \mathcal{B}^{(n)}(0) = 0, \quad \mathcal{G}^{(n)}(0) = 0. \quad (4.17)$$

Finally, for the axion–vector coupling, we select $\beta = 25$. Given our above initial conditions, this entails an onset of strong-backreaction effects a few e -folds before the expected end of inflation on the slow-roll trajectory (see, e.g., Ref. [128]).

In Fig. 4.1, we illustrate the GEF results for this benchmark model. The left panel of this figure demonstrates the GEF's ability to capture the dynamical increase in gauge-field production during axion inflation. We show the evolution of the three-gauge field densities; $\rho_E = \mathcal{E}^{(0)}/2$ (red),

⁶Given the symmetries of $V(\phi)$ in Eq. (4.15), one may also exchange the sign for these initial conditions, $\varphi(0) \rightarrow -\varphi(0)$, $\dot{\varphi}(0) \rightarrow -\dot{\varphi}(0)$. For axion inflation, the change in sign of $\dot{\varphi}$ entails that the opposite gauge-field helicity is tachyonically enhanced by the rolling inflaton field (cf. chapter 3). With our choice of initial conditions, the negative gauge-field helicity mode is tachyonically enhanced.

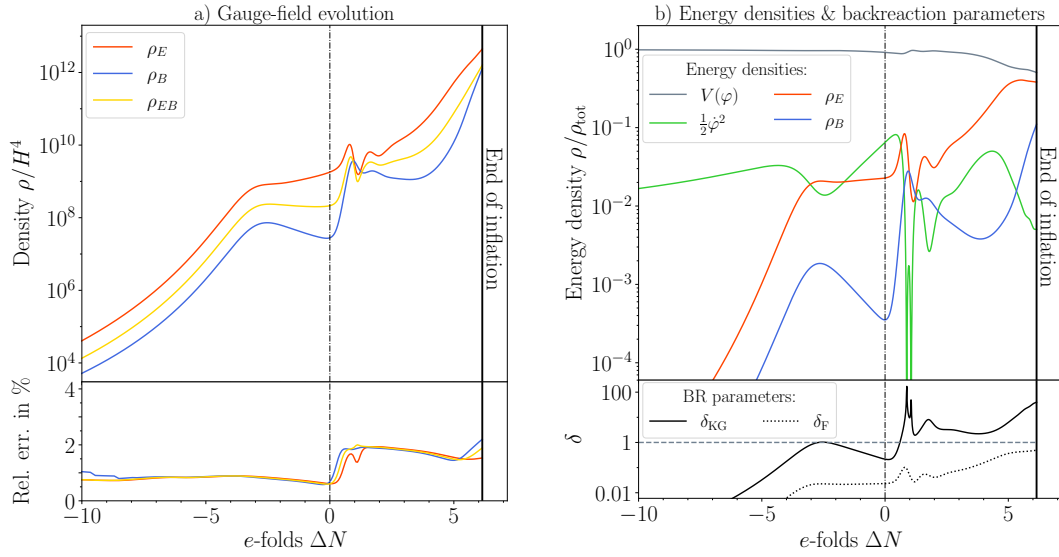


FIGURE 4.1: GEF results for axion inflation with an axion–vector coupling $\beta = 25$ and the inflaton potential in Eq. (4.15). **Panel (a):** The upper panel shows the evolution in e -folds of the electric energy density ρ_E (red), magnetic energy density ρ_B (blue), and the Chern–Pontryagin density ρ_{EB} (yellow), all normalized to H^4 . The lower panel shows the error estimate in Eq. (4.14) giving the deviation between the GEF result and the corresponding mode-by-mode solution in momentum space. **Panel (b):** The upper panel gives the relative contributions of the electric (red) and magnetic (blue) energy density to the total energy density. Also shown are the contributions of the inflaton’s kinetic (green) and potential (gray) energy densities to the total energy density. The lower panel shows the evolution of the backreaction parameters δ_{F} and δ_{KG} in Eq. (3.27).

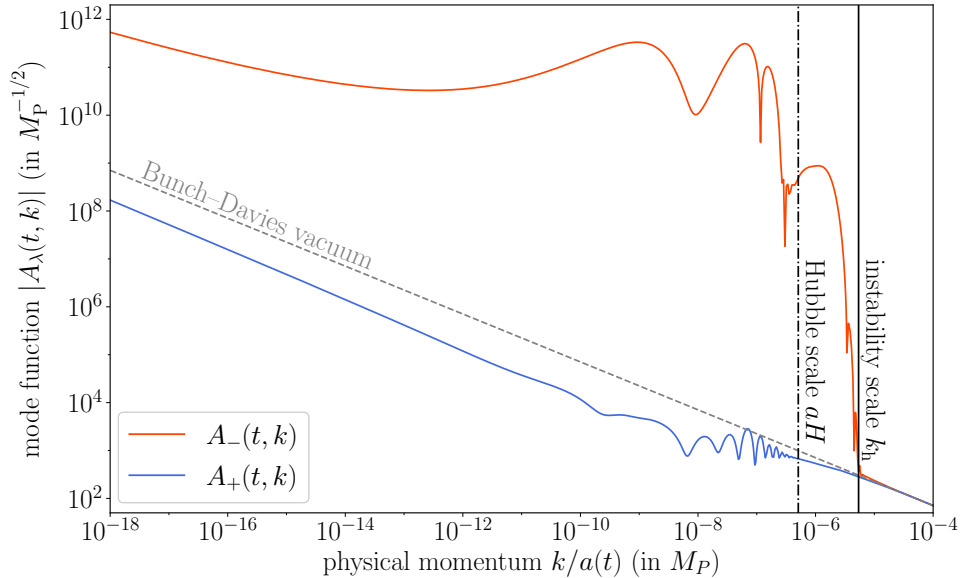


FIGURE 4.2: The gauge-field mode spectrum for the negative (red) and positive (blue) helicity mode at the end of inflation for $\beta = 25$ and $m = 6.16 \times 10^{-6} M_{\text{P}}$. The gray dashed line indicates the amplitude of the Bunch–Davies vacuum. Backreaction effects imprint themselves as oscillatory features below the instability scale k_{h} (black solid) for both helicity modes.

$\rho_B = \mathcal{B}^{(0)}/2$ (blue), which, respectively, correspond to the electric and magnetic energy density, and $\rho_{EB} = |\mathcal{G}^{(0)}|/2$ (yellow), the Chern–Pontryagin density. In the lower part of this plot, we also depict our estimate for the error on these quantities using the estimator in Eq. (4.14). This error consistently remains at the percent level, and clearly demonstrates the numerical control over our GEF solution, which persists even when oscillations appear in ρ_E , ρ_B , and ρ_{EB} .

In the right panel of Fig. 4.1, we instead demonstrate the GEF’s ability to capture the dynamical impact of gauge-field production. The upper panel shows the evolution of all energy densities in this model, normalized to the total energy density $\rho_{\text{tot}} = 3H^2 M_{\text{P}}^2$, while the lower panel shows the change in the backreaction parameters δ_{KG} and δ_{F} in Eq. (3.27). We see the system indeed transitions from weak to strong backreaction, as indicated by δ_{KG} reaching unity. In the strong-backreaction regime, the GEF result then exhibits all expected features (cf. Fig. 3.3): The onset of strong backreaction is followed by an increase in the duration of inflation (about 6 e -foldings), the system exhibits oscillatory features in ρ_E , ρ_B , and $\dot{\varphi}^2$, and the end of inflation is reached for $V(\varphi) \simeq \rho_E$.⁷

Finally, Fig. 4.2 depicts the spectrum of gauge-mode functions $A_\lambda(t, k)$ that one obtains based on the GEF solution shown in Fig. 4.1. The spectrum is evaluated at the end of inflation after the system has undergone several e -folds of strong backreaction. It is illustrative to compare this spectrum to the one shown in Fig. 3.1, which is unaffected by backreaction effects. Clearly, the spectrum in Fig. 4.2 exhibits novel signatures which we can attribute to the phase of strong backreaction. This is expected following our discussion in chapter 3: The mode spectrum carries a memory of the dynamical evolution of the inflationary system at earlier times [109]. Correspondingly, the violent oscillations of the inflaton’s velocity during the phase of strong backreaction imprint themselves on scales below the instability scale, k_{h} . This imprint is featured in both helicity modes, as the inflaton velocity briefly changes sign in the strong-backreaction regime (cf. Fig. 4.1 panel (b)). The GEF determination of the spectrum accurately captures these features as evidenced by the small numerical error at the end of inflation seen in Fig. 4.1 panel (a).

This exposition also demonstrates further advantages of the GEF that go beyond its relatively low numerical cost that we mentioned previously. With the GEF, we have access to a large dynamical range, which allows us to study the dynamics of inflation from the moment that the scales relevant for the CMB leave the horizon until its end. In particular, this also means that we can access a wide range in wavenumber for the gauge-field mode spectrum spanning multiple orders of magnitude (see Fig. 4.2). This is particularly relevant for our later applications of the GEF in part III of this thesis, where we use this spectrum of gauge-mode functions to compute the imprint of gauge fields on the power spectra of inflationary perturbations.

4.3 Inhomogeneous backreaction and the GEF

To have a complete perspective regarding the GEF’s abilities, we next want to highlight its main limitation. As we have alluded to in the previous chapter, several recent and independent lattice simulations (e.g., Refs. [378–380]) have demonstrated that the dynamics of axion inflation in the regime of strong backreaction are more complex than previously anticipated. In particular, these simulations show that inhomogeneities in the inflaton field are abundantly sourced in the regime of strong backreaction, which then proceed to alter the dynamics of the system. At present, neither the GEF nor the mode-by-mode approach to studying axion inflation can account for these effects of inflaton inhomogeneities. For simplicity, we make the distinction that lattice simulations are

⁷We define the parameter ΔN such that it reaches zero for the expected end of slow-roll inflation, approximately 60 e -foldings after initialization.

able to resolve *inhomogeneous backreaction*, while the GEF and mode-by-mode approach only describe *homogeneous backreaction*.

To understand the difference, let us briefly explore how gauge fields dynamically source inflaton inhomogeneities, and what interactions allow them to backreact on gauge-field production. Inflaton inhomogeneities, $\delta\phi(t, \mathbf{x}) = \phi(t, \mathbf{x}) - \varphi(t)$, may be sourced by the gauge field in the same way as the inflationary gauge field backreacts on the homogeneous inflaton field, φ . This can be seen via the perturbed Klein–Gordon equation for $\delta\phi$ (see, e.g., Ref. [56]):

$$\delta\ddot{\phi} + 3H\delta\dot{\phi} - \frac{1}{a^2}\Delta\delta\phi + V_{,\phi\phi}(\varphi)\delta\phi = \frac{\beta}{M_{\text{P}}}(\mathbf{E} \cdot \mathbf{B} - \langle \mathbf{E} \cdot \mathbf{B} \rangle) \quad (4.18)$$

where, as always, $\Delta = \nabla^2$ denotes the spatial Laplacian. Evidently, the right-hand side gauge-field term acts as an effective source for $\delta\phi$.

Once these inhomogeneities are sourced, they further affect gauge-field production through their appearance in Maxwell’s equations (3.12), in particular Ampère’s law (3.12b), and Gauss’ law (3.12a). Accounting for inflaton inhomogeneities, Ampère’s law becomes

$$\dot{\mathbf{E}} + 2H\mathbf{E} - \frac{1}{a} \text{rot } \mathbf{B} + \frac{\beta}{M_{\text{P}}} \left((\dot{\varphi} + \delta\dot{\phi})\mathbf{B} + \frac{1}{a}\nabla\delta\phi \times \mathbf{E} \right) = 0 \quad (4.19)$$

The important novelty in this equation is the term $(\beta/M_{\text{P}})(\nabla\delta\phi/a) \times \mathbf{E}$, which allows for an exchange of momenta between the electric field and the inflaton field. Another implication of inflaton inhomogeneities is their appearance in Gauss’ law, which becomes

$$\text{div } \mathbf{E} + \frac{\beta}{M_{\text{P}}} \mathbf{B} \cdot \nabla\delta\phi = 0. \quad (4.20)$$

The new term, $(\beta/M_{\text{P}})\mathbf{B} \cdot \nabla\delta\phi$, implies that the vector field $\mathbf{A}$ no longer describes only two transverse polarizations, but that there is an additional longitudinal polarization which can be sourced by the inflaton gradients [378, 379].

Next, let us discuss the dynamical implication of inhomogeneous backreaction in practice. To this end, a useful reference are the results obtained by Ref. [378], who studied both homogeneous and inhomogeneous backreaction on the lattice and directly compared these results to existing GEF and mode-by-mode data. Figure 4.3 summarizes the results of their comparison, which we take directly from Ref. [378]. Each column in Fig. 4.3 depicts the dynamics of axion inflation for different values of the axion–vector coupling, $\beta = 15, 18, 20$ (labelled α_{Λ} in Fig. 4.3) with the same inflaton potential as in Eq. (4.15). The first row shows the evolution of the inflaton’s potential (gray), kinetic (red), and gradient (blue) energy density alongside the gauge-field energy density (magenta). The second row instead depicts the evolution of the instability parameter ξ (black). Results are shown for the two kinds of lattice simulations: those assuming homogeneous backreaction (dashed) and those for inhomogeneous backreaction (solid). The lower panel also depicts results from the GEF (green, solid) and mode-by-mode approach (magenta, dashed).

Most relevant for us is the comparison to the GEF results in the second row of Fig. 4.3. First, we can see that the GEF shows excellent agreement with the lattice results for homogeneous backreaction, as well as the iterative mode-by-mode procedure. Moreover, the GEF also agrees with the inhomogeneous lattice study before the onset of backreaction effects, and approximately one e -fold leading into the stage of backreaction. This is important to note, as it indicates that the GEF accurately captures the onset of strong backreaction. Thereafter, the GEF only qualitatively reproduces the dynamics of inhomogeneous backreaction. This is unmistakably attributed to the growth in axion gradient energy in the upper row of Fig. 4.3. Inhomogeneous backreaction still

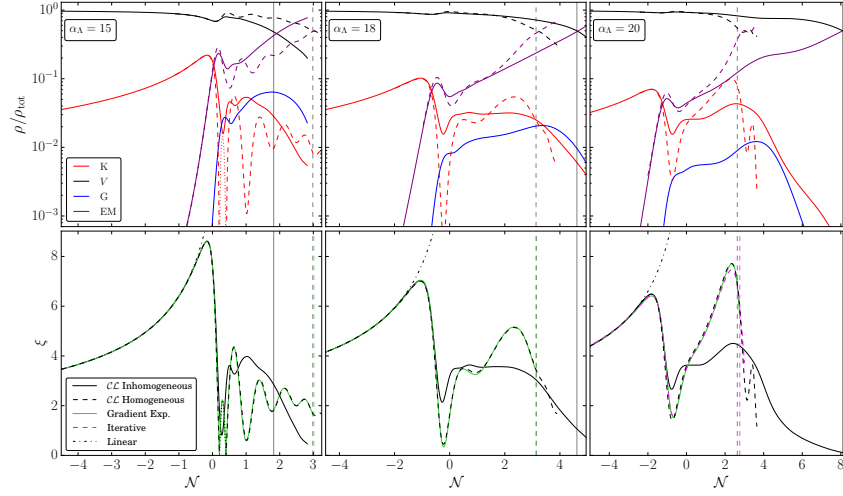


FIGURE 4.3: A comparison between homogeneous and inhomogeneous backreaction for the potential in Eq. (4.15). Each column corresponds to a different axion–vector coupling $\beta = 15$ (left), 18 (middle) and 20 (right) (labelled α_A in this figure). The first row shows the evolution of different energy densities: the inflaton’s potential (gray), kinetic (red), and gradient (blue) energy density and the electromagnetic energy density (magenta). The second row shows the evolution of the instability parameter ξ . Solid lines correspond to full lattice simulations including inhomogeneous backreaction, while the dashed lines indicate lattice simulations for homogeneous backreaction. Green lines show GEF results for homogeneous backreaction. The time parameter, $\mathcal{N}$ gives the number of e -folds starting from the expected end of inflation based on pure slow-roll dynamics.

This figure is reproduced from Ref. [378].

extends the duration of inflation, but the oscillatory features in ξ and the gauge-field energy density are smoothed. The extended duration of inflation also differs between the homogeneous and inhomogeneous backreaction results (cf. the vertical dashed and solid lines in Fig. 4.3).

In light of these results, a relevant question we may be inclined to ask is why inhomogeneous backreaction cannot be incorporated into the GEF. The reason can be gleaned following our discussion in Sec 4.1. The GEF relies on the formulation of a closed set of differential equations for the quadratic gauge-field expectation values, $\langle E^2 \rangle$, $\langle \mathbf{E} \cdot \mathbf{B} \rangle$ etc., as given in Eq. (4.4). This construction is possible because Maxwell’s equations in the absence of inflaton inhomogeneities are fully linear in the quantum operators $\mathbf{E}$ and $\mathbf{B}$ (cf. Eq. (3.12)). This changes when we include the effects of inflaton gradients into Maxwell’s equations which now contains quadratic interactions between quantum operators, $\delta\phi$ and $\mathbf{E}$ (cf. Eq. (4.19)). To incorporate the effect of inflaton perturbations into the GEF therefore requires the inclusion of higher-order correlation functions between the electric field $\mathbf{E}$, the magnetic field $\mathbf{B}$, and $\delta\phi$. For instance, the evolution equation for $\langle E^2 \rangle$ depends on the three-point correlator $\langle \mathbf{E} \cdot (\nabla \delta\phi / a) \times \mathbf{E} \rangle$, whose evolution itself depends on four-point correlation functions of $\mathbf{E}$, $\mathbf{B}$, and $\delta\phi$, etc.

Notably, one attempt at capturing the effects of inhomogeneous backreaction with the GEF has been made in Ref. [372]. These authors included the effects of inflaton gradients into the GEF assuming that one may perturbatively expand in $\delta\phi$, such that the GEF equations account for the effects of three-point correlation functions between $\delta\phi$, $\mathbf{E}$ and $\mathbf{B}$ without the need for including higher-order correlation functions in these fields. However, this approach did not yield an improved agreement with lattice simulations of inhomogeneous backreaction. Indeed, it is possible that inhomogeneous backreaction is a fully nonperturbative dynamical regime of axion inflation, which cannot be described by a perturbative expansion in $\delta\phi$. It should however be noted that the adapted GEF presented in Ref. [372] only accounted for the sourcing of two transverse polarizations

of the gauge field, while the lattice results of Ref. [379] highlight that inflaton inhomogeneities are effective at also sourcing longitudinal gauge-field modes. This beckons the question if the perturbative method of Ref. [372] can be improved by also accounting for longitudinal gauge-field polarizations.⁸ Therefore, it is not conclusively clear if the GEF is fundamentally incapable of resolving the dynamics of inhomogeneous backreaction. However, at present, inhomogeneous backreaction remains a challenge for the GEF.

Finally, it is also important to stress that the effect of inhomogeneous backreaction which we just discussed has thus far only been confirmed for the case of axion inflation without the Schwinger effect. As we have briefly mentioned in chapter 3, the Schwinger effect can effectively drain energy from the electromagnetic background in axion inflation (see Eq. (3.35)), which can prohibit the onset of strong-backreaction dynamics [104, 128, 129, 134] (see also our later chapter on the Schwinger effect, chapter 6). This could imply that the Schwinger effect also prohibits the dynamical growth of gauge-field-sourced inflaton inhomogeneities, and thus prevents the onset of inhomogeneous backreaction in this case. Indeed, first lattice simulations which have incorporated this damping of gauge-field production during axion inflation [134] find no difference in the dynamics of axion inflation for homogeneous and inhomogeneous backreaction. However, this analysis has thus far only been carried out for particular benchmark models. Thus, given the current state of the literature, the inability of the GEF to capture inhomogeneous-backreaction effects appears to be less relevant for systems with effective Schwinger damping, as we will consider in chapter 6.

4.4 Outlook: the role of the GEF within axion inflation

As we have seen in this chapter, the GEF allows for an internally self-consistent determination of the dynamics of axion inflation including gauge-field backreaction, with its key advantages being its low numerical cost, and its ability to capture gauge-field production over a large range in e -folds. At the same time, we have also seen that the GEF inherently simplifies the equations of motions of axion inflation, in that it relies on the assumption of negligible inflaton inhomogeneities, while lattice results indicate that these effects can be important in the strong-backreaction regime. Given that the GEF is the main tool which we rely on in this thesis, we now want to close this chapter by considering how its limitations and advantages with respect to lattice simulations delimit its specific role in the study of axion inflation and its regime of applicability.

To start, let us clearly address the GEF's limitation: the fact that the GEF does not resolve the effects associated with inhomogeneous backreaction means that we may only use it to qualitatively describe gauge-field production in the strong-backreaction regime. Important to emphasize is that the GEF is not entirely unsuitable for studying effects associated with strong backreaction. It accurately captures the onset of strong backreaction, and reproduces features that are also present for inhomogeneous backreaction, e.g., an extended duration of inflation. Nevertheless, this means that, when using the GEF to study phenomenological implications of gauge-field production, any effects that one observes which are associated with strong-backreaction dynamics are likely only qualitatively accurate. This is particularly relevant in the part III of this thesis, where we use the GEF to study how gauge-field production imprints itself in the spectrum of inflationary perturbations. For these analyses, we need to be keenly aware of our limitations, and our conclusions may require further scrutiny by follow-up lattice studies, which can make more quantitatively accurate

⁸Relevant in this context is the analysis of Ref. [392], which recently presented a GEF version which describes the evolution of both transverse and longitudinal gauge-field polarizations to study the production of massive vector fields during inflation.

predictions once strong-backreaction effects are relevant. In particular, in these subsequent chapters we will emphasize this point appropriately.

Meanwhile, the GEF's key advantages, i.e., its low cost and large dynamical range, also means it excels in several aspects where lattice simulations are limited. In particular, several analyses presented in this thesis make use of these advantages, and we can already anticipate them here to highlight the relevance of the GEF technique.

Firstly, the numerical efficiency of the GEF means that it is ideally suited for studies that aim at exploring vast regions of the parameter space for axion inflation models. This is in contrast to the relatively costly lattice simulations, which, at the moment, are aimed at select benchmark scenarios (see Refs. [376–381]). As an example, in chapter 9, the GEF allows us to perform a comprehensive parameter space analysis concerning the production and detectability of gravitational waves from axion inflation. Such searches may then be seen as establishing a *GEF benchmark* which may guide follow-up lattice studies.

Secondly, the GEF's ability to simulate inflationary dynamics over many e -foldings means that it can be used to study phenomenological effects of gauge-field production on physical scales spanning several orders of magnitude in wavenumber. For instance, in chapter 8, this allows us to identify unexpected effects of late-time gauge-field production on large-scale features in the scalar power spectrum. Such effects are harder to identify when studying inflationary dynamics on a lattice, as these are limited in their dynamical range by a finite lattice size and spacing (see, e.g., Ref. [395, 396]).

Moreover, lattice methods may rely on the output of homogeneous-backreaction techniques like the GEF for calibration purposes and for obtaining initial conditions for their simulations (see, e.g., Ref. [378, 379]). We should therefore view the GEF and the lattice as complementary and not adversarial tools for studying axion inflation. Each of these techniques brings with it clear advantages, regimes of applicability, and limitations.

Moving forward, in the next chapter we will consider a particularly unique application of the GEF, in which we directly initialize the GEF in the regime of strong gauge-field backreaction to study if gauge-field production can provide a stable source of friction in axion inflation, as had originally been proposed by ANBER AND SORBO in Ref. [55]. This application is entirely unique for the GEF, as it takes advantage of the GEF's particular perspective on the equations of motion for axion inflation, which are unlike the usual momentum-space description of axion inflation as we have presented in chapter 3.

Beyond this application of the GEF, in chapter 6, we extend the GEF to incorporate Schwinger pair production into axion inflation, including the damping effects it has on gauge-field production. Thereafter, in chapters 8 and 9, we use the GEF-generated background dynamics as the basis for computing scalar and tensor perturbations in axion inflation.

Chapter 5

The stability of the Anber–Sorbo solution

This chapter is based on Ref. [1], with a partial input from Ref. [6], in particular in Secs. 5.2 and 5.4. The detailed contributions of the author to these articles are outlined at the beginning of this thesis.

In the previous chapter, we have developed a basic understanding of the gradient-expansion formalism (GEF), including its construction, its numerical application, and its advantages and disadvantages. As part of this presentation, we focussed on the GEF’s main purpose that we rely on for the later parts of this thesis, namely, using it as a tool for numerically computing inflationary dynamics for axion inflation starting from slow-roll initial conditions. However, the unique formulation of the equations of motion of axion inflation that the GEF provides also proves to be a useful basis for developing novel diagnostic tools that we can use to improve our understanding of gauge-field backreaction effects. In this chapter, we demonstrate this application of the GEF by using it to study the stability of the so-called *Anber–Sorbo solution* of axion inflation [55] — a possible stable attractor solution in the strong-backreaction regime of axion inflation where its dynamics could be studied fully analytically.

More specifically, we here use the GEF as a basis for developing two diagnostic tools which permit us to directly examine the stability of the Anber–Sorbo (AS) solution. One of these tools is what we refer to as the *linearized gradient-expansion formalism* (or LGEF), which reduces the study of nonlinear backreaction effects in axion inflation to that of a simple set of linear equations which can be solved using conventional linear-algebra techniques. This tool is combined with the conventional GEF we have described in the previous chapter, which we initialize directly in the AS solution to study how perturbations to the AS solution dynamically evolve over time. In combination, these two techniques allow us to solidify prior claims brought forth in Ref. [131] that the AS solution is unstable. However, beyond confirming these results, we also can identify a region in the phase space of axion inflation where the AS solution may indeed be stable.

We begin this discussion in Sec. 5.1 by reviewing the AS solution and the current state of the literature regarding its stability. In Sec. 5.2, we then develop the tools that we need to analyze its stability, including the LGEF. Section 5.3 is where we apply these GEF-based tools to perform a comprehensive stability analysis of the AS solution, while in Sec. 5.4 we highlight what other insights the applications of the LGEF can give into the dynamical structure of axion inflation. Finally, we end our discussion in Sec. 5.5 with a summary of our results and a conclusion.

Notation: In the following, we work with the model of axion inflation described in Sec. 3.2. This means, the dynamics of axion inflation are given by the Klein–Gordon equation (3.11) for the background inflaton field, $\varphi = \langle \phi \rangle$, Maxwell’s equations (3.12) for the electric- and magnetic field operators, $\mathbf{E}$ and $\mathbf{B}$, and the Friedmann equation (3.13) for the Hubble rate H . Furthermore, V denotes the inflaton potential, $V_{,\phi}$ its gradient, and β is the dimensionless strength of the axion–gauge-field coupling.

5.1 What is the Anber–Sorbo solution?

The AS solution was first presented in Ref. [55] as an alternative realization of slow-roll inflation during axion inflation which can be uniquely realized in the regime of strong gauge-field backreaction. Unlike the conventional slow-roll attractor, where the inflaton’s slow-roll with a nearly constant velocity is sustained by Hubble friction, $3H\dot{\phi} \simeq -V_{,\phi}$ (cf. Eq. (1.29)), the AS solution suggests that the inflaton’s velocity may be kept constant by balancing the inflaton’s potential gradient against strong gauge-field friction, i.e.,

$$V_{,\phi}(\varphi) \simeq \frac{\beta}{M_{\text{P}}} \langle \mathbf{E} \cdot \mathbf{B} \rangle, \quad (5.1)$$

If this balance is consistently maintained, it would allow for a period of inflation on a steep potential, where the conventional slow-roll solution cannot be realized. Moreover, on this new attractor, the otherwise complex and nonlinear dynamics of strong backreaction could be described using entirely analytical methods.

The original assumption that lead to the proposal of the AS attractor in Ref. [55] is that the instability parameter $\xi = \beta\dot{\phi}/(2HM_{\text{P}})$ (cf. Eq. (3.20)), i.e., the parameter regulating gauge-field production, can be kept constant in the regime of strong gauge-field backreaction. Indeed, under idealizing assumptions of a constant Hubble rate during inflation, the equations of motion for axion inflation do permit a steady-state solution with a constant instability parameter. We have previously seen this in chapter 3 when discussing the dynamics of axion inflation in the weak-backreaction regime, where we noted that constant ξ and constant H allow one to have a perfectly constant gauge-field friction term $\langle \mathbf{E} \cdot \mathbf{B} \rangle$. These simplifying assumptions then allow to establish the perfect balance between gauge-field friction and the inflaton’s potential gradient as given in Eq. (5.1).

However, just because this balance can be achieved under idealized assumptions that does not mean it is also realized in a realistic model of axion inflation where H slowly varies as a function of time. Moreover, even when making these simplifications, the mere existence of a steady-state solution does not imply that it is an attractor. Thus, let us review the current evidence against the idea that the AS solution is a stable attractor.

The first evidence is empirical: There are no numerical simulations of axion inflation which reach the supposed AS attractor when evolving the dynamics of axion inflation from the weak to the strong-backreaction regime. Instead, it is observed that strong backreaction immediately leads to nonlinearities (see for instance, Refs. [106–109] or Figs. 3.3, 4.1, and 4.3). Indeed, these observations align with the conclusions of Ref. [109], where it was demonstrated that gauge-field friction reacts retardedly to changes in the instability parameter ξ . This retarded response implies that it is difficult to achieve the delicate balance in Eq. (5.1) needed to ensure the AS solution persists. Meanwhile, this numerical evidence leaves open the possibility that the AS solution’s basin of attraction is merely small, such that simulations starting from slow-roll initial conditions are prone to miss it. To definitively demonstrate the AS solution’s instability, one rather should study its stability by directly starting one’s simulation from the AS solution itself.

The second set of evidence suggesting that the AS solution is unstable are the analytical results of Ref. [131]. Its authors directly studied the stability of the AS solution under the idealizing assumptions of a constant Hubble rate and a constant instability parameter, ξ . To this end, they constructed a system of linearly perturbed equations of motion for axion inflation around the AS solution itself by accounting for perturbations to both the instability parameter and to the gauge-field mode functions $A_{\lambda}(t, k)$. By solving their perturbed equations of motion, they identify that perturbations to the instability parameter, $\delta\xi$, necessarily grow as a function of time, signaling the

eventual instability of the AS solution even when assuming constant H .

The results of Ref. [131] are a clear improvement over the purely empirical evidence from numerical studies, as its authors assess the stability of the AS solution directly, instead of attempting to reach it dynamically. However, their analysis still leaves further open questions. Firstly, to demonstrate the instability of the AS solution, the authors of Ref. [131] relied on a series of approximations to solve their system of perturbed equations. Moreover, their linear perturbation analysis does not reveal what dynamics follow after the AS solution has become destabilized. Finally, their analytical techniques do not give insight into the stability of the AS solution beyond the assumption of a constant Hubble rate. After all, it is in principle conceivable, albeit unlikely, that mild variations in H may stabilize the AS solution after all.

In what follows, our aim is to address these questions using the GEF. To do so, we rely on two complementary approaches based on the GEF which together allow for a comprehensive stability analysis of the AS solution. The first tool, the linearized gradient-expansion formalism, is designed to improve upon the linear stability analysis made in Ref. [131]. Matching the assumption of a constant Hubble rate H made in this work, we derive a linear set of ordinary differential equations describing perturbations around the AS solution by relying on the GEF formulation of axion inflation. This way, we can easily demonstrate the instability of the AS solution by identifying unstable eigenmodes in this linearized system without relying on the approximations made in Ref. [131]. The second tool at our disposal is the full nonlinear GEF itself. Its initial conditions can be tuned to the AS solution to numerically study how deviations from the AS solution evolve over time. This second approach permits us to go beyond linear perturbation theory, and can also capture the late-time behavior of axion inflation once the AS solution has decayed. Moreover, the full GEF can be applied both under the assumption of constant H , and for varying H , and therefore allows further insight into the AS solution's stability in more realistic models of axion inflation.

5.2 Stability analysis with the GEF

Let us develop the set of tools that we shall rely on for our stability analysis. As outlined in the previous section, the AS solution we intend to study is derived under the assumption that the Hubble rate during inflation is exactly constant. For the majority of our stability analysis, we will also rely on this assumption, in line with the approach taken in Ref. [131]. Only in the last part of this analysis do we also account for variations of the Hubble rate. Note also that our analysis entirely neglects the impact of inflaton inhomogeneities, whose effects were also disregarded in the derivation of the original AS solution.¹

In the following, we will work with a particular toy model of axion inflation for which the AS solution can be realized *exactly* by choosing appropriate initial conditions. This allows us to give a clean, reproducible definition of the AS solution, which we can rely on for our stability analysis. In addition to a constant Hubble rate, our toy model also assumes a particularly simple inflaton potential

$$V(\phi) = V_0 - \kappa(\phi - \phi_0), \quad (5.2)$$

with both κ and the potential height V_0 being constant. This may be viewed as the first order Taylor expansion of an arbitrary potential around some fixed field value for the inflaton, ϕ_0 . The idea behind the AS solution is that gauge-field friction balances the gradient $V_{,\phi} = -\kappa$ exactly, thus higher-order terms in the Taylor expansion of $V(\phi)$ are irrelevant.

¹It is at present not known if the full system of axion inflation including inhomogeneous backreaction even permits a dynamical configuration analogous to the AS solution. The question regarding the AS solution's stability is in this sense only properly phrased assuming inhomogeneous-backreaction effects are absent, even if they are likely present in a fully realistic model of axion inflation. We also further comment on this in Sec. 5.4.

Let us then consider the set of dynamical equations describing this toy model. The assumption $H = \text{const.}$ means that the dynamics of axion inflation are entirely defined by the Klein–Gordon equation (3.11), and the mode equation for the gauge-field mode functions $A_\lambda(\eta, k)$, Eq. (3.19), while we simply neglect the dynamical evolution of H implied by the Friedmann equation. This allows us to conveniently rewrite the Klein–Gordon equation as a dimensionless first-order differential equation for the instability parameter ξ :

$$\frac{d\xi(N)}{dN} + 3\xi(N) - \frac{bv}{2} = -\frac{b^2}{2}g_0(N). \quad (5.3)$$

Here, we work with the number of e -folds as a time parameter, $N = \ln a = Ht$, and we have defined the dimensionless variables

$$v \equiv -\frac{\kappa}{H^3}, \quad b \equiv \frac{\beta}{M_{\text{P}}}H. \quad (5.4)$$

The function $g_0(N)$ is defined as a dimensionless version of the gauge-field friction term, $g_0 \equiv -\langle \mathbf{E} \cdot \mathbf{B} \rangle / H^4$ in Eq. (3.11). Equivalently, we can understand it as a rescaled version of the GEF variable $\mathcal{G}^{(0)}$, i.e., $g_0 = \mathcal{G}^{(0)} / H^4$ as defined in Eq. (4.1). The function $g_0(N)$ can therefore be readily expressed in terms of $A_\lambda(\eta, k)$ as given in Eq. (4.5). For constant H , these solve (cf. Eq. (3.19))

$$A_\lambda''(\eta, k) + \left(k^2 + \frac{2\lambda k \xi}{\eta}\right) A_\lambda(\eta, k) = 0, \quad (5.5)$$

where we have used the exact relationship $\eta = -1/(aH)$ in de Sitter space.

It is then easy to demonstrate that the coupled set of differential equations (5.3) and (5.5) allow for an exact solution in terms of a constant instability parameter. Assuming $\xi(N) = \bar{\xi} = \text{const.}$, one may solve the mode equation (5.5) exactly in terms of Whittaker functions (cf. chapter 3 and appendix A):

$$A_\lambda(\eta, k) = \frac{e^{\lambda\pi\bar{\xi}/2}}{\sqrt{2k}} W_{-i\lambda\bar{\xi}, 1/2}(2ik\eta). \quad (5.6)$$

This way, one can then determine $g_0(N)$ as an exact, time-independent, function of $\bar{\xi}$ by plugging Eq. (5.6) into the integral expression for $\mathcal{G}^{(0)}$ in Eq. (4.5). The expression is

$$\bar{g}_0(\bar{\xi}) = \frac{1}{4\pi^2} \int_0^{2|\bar{\xi}|} dx \sum_{\lambda=\pm} \lambda x^2 e^{\lambda\pi\bar{\xi}} \text{Re} \left[W_{i\lambda\bar{\xi}, \frac{1}{2}}(2ix) W_{1-i\lambda\bar{\xi}, \frac{1}{2}}(-2ix) \right], \quad (5.7)$$

where $x = -k\eta$, and we used the properties of Whittaker functions given in appendix A. Therefore, taking Eq. (5.7) and returning to the dimensionless Klein–Gordon equation in Eq. (5.3), we find that $\bar{\xi} = \text{const.}$ solves this equation exactly if it satisfies the condition

$$6\bar{\xi} - bv + b^2\bar{g}_0(\bar{\xi}) = 0, \quad (5.8)$$

This then defines a manifold of possible constant- ξ solutions to Eq. (5.3), i.e., $\bar{\xi} = \bar{\xi}(v, b)$.

We can qualitatively classify such solutions in two categories. For the first kind, gauge-field backreaction is weak, i.e., $\bar{g}_0(\bar{\xi})$ in Eq. (5.7) is small compared to the Hubble friction term $3\bar{\xi}$. Such solutions correspond to conventional slow-roll dynamics which axion inflation can realize in the weak-backreaction regime (cf. chapter 3). The second kind of solutions exist for strong-backreaction effects and correspond to the AS solution in our toy model. The two regimes can also be distinguished using the backreaction parameters for axion inflation defined in Eq. (3.27). For

the AS solution, we qualitatively require

$$\delta_{\text{KG}} \gg 1, \quad \delta_{\text{F}} \ll 1, \quad (5.9)$$

i.e., a strong effect of gauge-field friction on the dynamics of the inflaton field, but a negligible contribution of gauge-field energy density in the Friedmann equation.² In contrast, the conventional slow-roll attractor occurs for $\delta_{\text{KG}} \ll 1$ and $\delta_{\text{F}} \ll 1$.

For the purposes of our stability analysis, we now want to consider the evolution of perturbations away from a solution $\bar{\xi} = \text{const.}$ to Eq. (5.8). That is, we study the time-dependent evolution of the instability parameter ξ for the Ansatz

$$\xi(N) = \bar{\xi} + \delta\xi(N). \quad (5.10)$$

As we know from our discussion in chapter 3, once ξ is fully time-dependent, we may no longer assume that the friction term $g_0(N)$ in Eq. (5.3) is a time-independent function of $\bar{\xi}$. Instead, we need to determine $g_0(N)$ dynamically. To describe these dynamics, it is then useful to change to a GEF formulation of our toy model, in the spirit of the construction that we outlined in chapter 4.

5.2.1 The toy-model GEF

From the perspective of the GEF, the friction term g_0 on the right-hand side of Eq. (5.3) is merely one of a set of functions describing the background gauge-field configuration in our toy model. These functions are the dimensionless versions of the GEF variables $\mathcal{E}^{(n)}$, $\mathcal{G}^{(n)}$ and $\mathcal{B}^{(n)}$ which we defined in Eq. (4.1):

$$e_n(N) \equiv \frac{1}{H^{n+4}} \mathcal{E}^{(n)}(N) = \frac{1}{a^n H^{n+4}} \langle \mathbf{E} \cdot \text{rot}^n \mathbf{E} \rangle, \quad (5.11a)$$

$$g_n(N) \equiv \frac{1}{H^{n+4}} \mathcal{G}^{(n)}(N) = -\frac{1}{2a^n H^{n+4}} \langle \mathbf{E} \cdot \text{rot}^n \mathbf{B} + \mathbf{B} \cdot \text{rot}^n \mathbf{E} \rangle, \quad (5.11b)$$

$$b_n(N) \equiv \frac{1}{H^{n+4}} \mathcal{B}^{(n)}(N) = \frac{1}{a^n H^{n+4}} \langle \mathbf{B} \cdot \text{rot}^n \mathbf{B} \rangle. \quad (5.11c)$$

Analogous to the conventional GEF equations in Eq. (4.4), their evolution in our toy model is governed by a system of coupled ODE's, namely

$$\frac{de_n}{dN} + (n+4)e_n - 4\xi g_n + 2g_{n+1} = \frac{S_{\mathcal{E}}^n}{H^{n+5}}, \quad (5.12a)$$

$$\frac{dg_n}{dN} + (n+4)g_n - 2\xi b_n + b_{n+1} - e_{n+1} = \frac{S_{\mathcal{G}}^n}{H^{n+5}}, \quad (5.12b)$$

$$\frac{db_n}{dN} + (n+4)b_n - 2g_{n+1} = \frac{S_{\mathcal{B}}^n}{H^{n+5}}, \quad (5.12c)$$

with the boundary terms, $S_{\mathcal{X}}^n$ given in Eq. (4.9). We may of course truncate these equations at a finite order n_{cut} following the prescription in Eq. (4.13).

²The condition $\delta_{\text{F}} \ll 1$ is needed to ensure that our toy-model assumption $H = \text{const.}$ approximately reflects the more realistic model of axion inflation where H may slowly vary during inflation.

If the instability parameter ξ is perfectly tuned to the Anber–Sorbo solution, i.e., $\xi(N) = \bar{\xi} = \text{const.}$, all functions $e_n(N)$, $g_n(N)$, $b_n(N)$ are exact functions of $\bar{\xi}$:

$$\bar{e}_n(\bar{\xi}) = \frac{1}{4\pi^2} \int_0^{2|\bar{\xi}|} dx \sum_{\lambda=\pm} \lambda^n x^{n+1} e^{\lambda\pi\bar{\xi}} \left| (x - \lambda\bar{\xi}) W_{-i\lambda\bar{\xi}, \frac{1}{2}}(-2ix) - i W_{1-i\lambda\bar{\xi}, \frac{1}{2}}(-2ix) \right|^2, \quad (5.13a)$$

$$\bar{g}_n(\bar{\xi}) = \frac{1}{4\pi^2} \int_0^{2|\bar{\xi}|} dx \sum_{\lambda=\pm} \lambda^{n+1} x^{n+2} e^{\lambda\pi\bar{\xi}} \text{Re} \left[W_{i\lambda\bar{\xi}, \frac{1}{2}}(2ix) W_{1-i\lambda\bar{\xi}, \frac{1}{2}}(-2ix) \right], \quad (5.13b)$$

$$\bar{b}_n(\bar{\xi}) = \frac{1}{4\pi^2} \int_0^{2|\bar{\xi}|} dx \sum_{\lambda=\pm} \lambda^n x^{n+3} e^{\lambda\pi\bar{\xi}} \left| W_{-i\lambda\bar{\xi}, \frac{1}{2}}(-2ix) \right|^2. \quad (5.13c)$$

These expressions are analogous to Eq. (5.7) and follow from Eqs. (4.5), (5.6), and (5.11).

For a fixed value of $\bar{\xi} = \text{const.}$, one can therefore exactly (up to numerical precision) tune the system of functions in Eq. (5.11) to the AS solution by setting the initial conditions

$$e_n(0) = \bar{e}_n(\bar{\xi}), \quad g_n(0) = \bar{g}_n(\bar{\xi}), \quad b_n(0) = \bar{b}_n(\bar{\xi}), \quad (5.14)$$

at a given initial time defined as $N = 0$. By perturbing these initial conditions, for instance adding a small initial offset to the instability parameter, $\xi(0) = \bar{\xi} + \delta\xi(0)$, the GEF equations in Eq. (5.12) can then be numerically solved to study the stability of the AS solution. If the AS solution is indeed unstable, perturbations from AS will grow and eventually lead to the onset of fully nonlinear dynamics.

5.2.2 The linearized gradient-expansion formalism

We can develop a different perspective on the stability of the AS solution by specializing to the regime where deviations from the exact AS solution are perturbatively small. That is, we study the time evolution of the following set of perturbation variables

$$\delta\xi \equiv \xi - \bar{\xi}, \quad \delta e_n \equiv e_n - \bar{e}_n(\bar{\xi}), \quad \delta g_n \equiv g_n - \bar{g}_n(\bar{\xi}), \quad \delta b_n \equiv b_n - \bar{b}_n(\bar{\xi}), \quad (5.15)$$

for a fixed constant value $\bar{\xi}$.

Expanding the GEF equations, Eq. (5.12), and the equation of motion for ξ , Eq. (5.3), to linear order in these variables yields the following set of linear coupled ODEs:

$$\frac{d\delta\xi}{dN} + 3\delta\xi = -\frac{b^2}{2}\delta g_0, \quad (5.16a)$$

$$\frac{d\delta e_n}{dN} + (n+4)\delta e_n - 4\bar{\xi}\delta g_n - 4\bar{g}_n(\bar{\xi})\delta\xi + 2\delta g_{n+1} = \delta S_n^{(e)}, \quad (5.16b)$$

$$\frac{d\delta g_n}{dN} + (n+4)\delta g_n - 2\bar{\xi}\delta b_n - 2\bar{b}_n(\bar{\xi})\delta\xi + \delta b_{n+1} - \delta e_{n+1} = \delta S_n^{(g)}, \quad (5.16c)$$

$$\frac{d\delta b_n}{dN} + (n+4)\delta b_n - 2\delta g_{n+1} = \delta S_n^{(b)}. \quad (5.16d)$$

Here, the right-hand side source terms are derived from Eq. (4.9) and are given by

$$\delta S_n^{(e)} = \frac{(2\bar{\xi})^{n+4}}{4\pi^2} \sum_{\lambda=\pm} \lambda^n \left\{ \frac{\mathcal{E}_\lambda(\bar{\xi})}{\bar{\xi}} \left[(n+4)\delta\xi + \frac{d\delta\xi}{dN} \right] + \frac{d\mathcal{E}_\lambda(\bar{\xi})}{d\bar{\xi}} \delta\xi \right\}, \quad (5.17a)$$

$$\delta S_n^{(g)} = \frac{(2\bar{\xi})^{n+4}}{4\pi^2} \sum_{\lambda=\pm} \lambda^{n+1} \left\{ \frac{\mathcal{G}_\lambda(\bar{\xi})}{\bar{\xi}} \left[(n+4)\delta\xi + \frac{d\delta\xi}{dN} \right] + \frac{d\mathcal{G}_\lambda(\bar{\xi})}{d\bar{\xi}} \delta\xi \right\}, \quad (5.17b)$$

$$\delta S_n^{(b)} = \frac{(2\bar{\xi})^{n+4}}{4\pi^2} \sum_{\lambda=\pm} \lambda^n \left\{ \frac{\mathcal{B}_\lambda(\bar{\xi})}{\bar{\xi}} \left[(n+4)\delta\xi + \frac{d\delta\xi}{dN} \right] + \frac{d\mathcal{B}_\lambda(\bar{\xi})}{d\bar{\xi}} \delta\xi \right\}, \quad (5.17c)$$

with the terms $\mathcal{E}_\lambda$, $\mathcal{G}_\lambda$, and $\mathcal{B}_\lambda$ defined in Eq. (4.10). Just as for the conventional GEF, this linear system of equations can be truncated to obtain a finite set of differential equations. A simple procedure to do so is to assume that, for a given value n_{cut} , all functions e_n , b_n , and g_n are exactly given by their unperturbed value, such that

$$\delta e_n = \delta g_n = \delta b_n = 0 \quad \text{for} \quad n > n_{\text{cut}}. \quad (5.18)$$

With this truncation procedure we then have a finite set of coupled linear ODEs for the perturbation variables in Eq. (5.15) with constant coefficients. These equations constitute the *linearized gradient-expansion formalism* (LGEF).

Schematically, the LGEF equations can be rewritten in the simple form

$$\frac{d}{dN} \vec{x}(N) = A \vec{x}(N), \quad (5.19)$$

where the vector $\vec{x}(N)$ is defined as

$$\vec{x} \equiv (\delta\xi, \delta e_0, \delta g_0, \delta b_0, \dots, \delta e_{n_{\text{cut}}}, \delta g_{n_{\text{cut}}}, \delta b_{n_{\text{cut}}})^T, \quad (5.20)$$

and where A is a time-independent matrix. Accordingly, general solutions to the LGEF equations, Eq. (5.16), can be found by determining the eigenvalues ζ_n for the matrix A . These solutions are simply given by

$$\vec{x}(N) = \sum_n e^{\zeta_n N} \vec{x}_0^{(n)}, \quad \text{with} \quad A \vec{x}_0^{(n)} = \zeta_n \vec{x}_0^{(n)}, \quad (5.21)$$

i.e., for $\vec{x}_0^{(n)}$ an eigenvector of A corresponding to the eigenvalue ζ_n . The eigenvalues of the matrix A constitute a spectrum of Lyapunov exponents. If any ζ_n in this spectrum has a positive real part, it signals that perturbations away from the AS solution continuously grow as a function of time, implying the instability of said solution. In particular, this means that one can resort to linear-algebra methods to study the stability of the AS solution instead of determining the full time-evolution of these perturbations, as we do when numerically solving the GEF equations in Eq. (5.12).³

5.3 Application to the Anber–Sorbo solution

Let us now apply both the conventional GEF and the LGEF to study the stability of the Anber–Sorbo solution for the toy model we just devised. At the end of this section, we also use these results as a basis for studying the effects of a varying Hubble rate.

³In practice, the truncation n_{cut} in Eq. (5.18) is chosen such that the eigenvalues are stable under a change in n_{cut} up to a desired numerical accuracy.

The two methods, GEF and LGEF, serve two distinct purposes in our analysis. The LGEF permits us to determine the full spectrum of Lyapunov exponents ζ_n and thus captures all possible evolutions for our perturbed system, independent of initial conditions, albeit only for as long as linear perturbation theory remains valid. Meanwhile, using the GEF, we can only study the particular evolution of our system for concrete initial perturbations, but we can trace this evolution beyond the validity of linear perturbation theory. Also, the GEF can even be applied when the Hubble rate is dynamical.

Before we begin our stability analysis, we should consider necessary conditions for the AS solution to be realized, i.e., we need to choose the free parameters of our model to guarantee that our system is in a regime of strong backreaction following Eq. (5.9). In our toy model, a constant- ξ solution is defined by two free parameters v and b , or, equivalently H , and β , which determine the constant value of ξ in Eq. (5.8) (referred to as ξ_0 in what follows). For our analysis, we work in terms of H , β and ξ_0 , and we focus on the particular region of the parameter space where Eq. (5.9) is satisfied alongside $\delta_F \delta_{KG} = 1$. Using this additional condition, we can derive an empirical expression relating H to β and ξ_0

$$\frac{H}{M_P} = 2 \times 10^{-7} \left(\frac{100}{\beta} \right)^{1/2} \exp(-2.85(\bar{\xi} - 7)) . \quad (5.22)$$

This condition follows from the expressions for δ_{KG} and δ_F for constant ξ in Eq. (3.32), while we fit the specific functional form in Eq. (5.22) to determine the numerical coefficients in this expression. If H , β and ξ_0 obey Eq. (5.22), we can safely assume that the system we are studying is configured to the AS solution. For definitiveness, we analyze the stability of the AS solution for $\xi_0 \in [5, 9]$ and $\beta \in [10^{1.5}, 10^{3.5}]$ with H given by Eq. (5.22).

5.3.1 Lyapunov exponents

It is most instructive to first discuss the results of our stability analysis using the LGEF method, as this already elegantly demonstrates our main results.

As explained in Sec. 5.1, the main point of comparison for the LGEF is the linear stability analysis of the AS solution performed in Ref. [131]. Therefore, to contextualize our subsequent results, let us first briefly elaborate on the main findings of this work. In their perturbative analysis, the authors of Ref. [131] studied linear perturbations to a constant instability parameter, ξ_0 alongside a perturbation δA_λ to the exact gauge-field mode function for constant ξ in Eq. (5.6). By making a power-law Ansatz for the inflaton perturbations $\delta \xi \propto e^{\zeta N}$, they determined a transcendental equation for the complex coefficient ζ in this Ansatz. This equation defines the spectrum of Lyapunov exponents of their linearized system (cf. Eq. (3.16) in our original article, Ref. [1]). Through a series of approximations, Ref. [131] derived an equation whose roots approximate the true Lyapunov exponents of their linear system. In our notation, this approximate equation reads [1]

$$R(\zeta) \equiv \left(\frac{\beta H}{M_P} \right)^2 \frac{315}{2^{15} \pi^2} \frac{e^{2\pi \bar{\xi}}}{\bar{\xi}^5} \frac{1}{(2\zeta - 1)(2\zeta + 7)} \left[\frac{\Gamma(2\zeta + 8)}{\Gamma(9)(8\bar{\xi})^{2\zeta - 1}} - 1 \right] = 1 . \quad (5.23)$$

Below, we shall compare the Lyapunov exponents determined from the LGEF to the approximate Lyapunov exponents given by the zeros of $|R(\zeta) - 1|$.

Let us first study the spectrum of Lyapunov exponents for a specific benchmark point, $\xi_0 = 7$ and $\beta = 10^{2.5}$, which is the central point in our region of interest. In Fig. 5.1, we show the results of the LGEF (red), alongside the roots of Eq. (5.23) (green circles). Clearly, both methods agree that there is a Lyapunov exponent with a positive real part. Therefore, both methods agree that

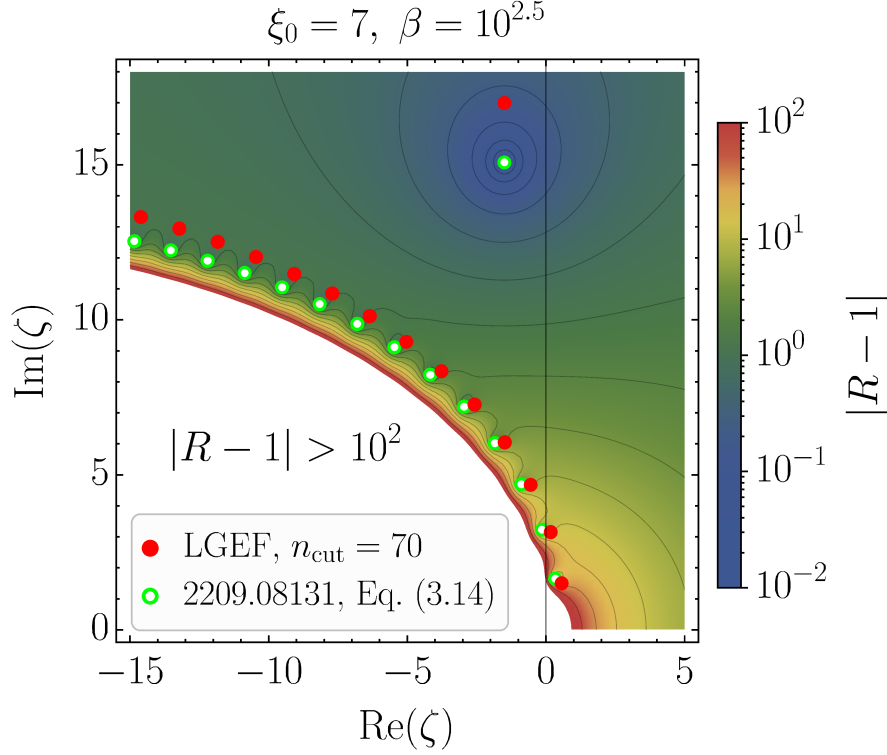


FIGURE 5.1: The eigenvalues of the LGEF system (for $n_{\text{cut}} = 70$) (red dots), and the solutions of Eq. (5.23) (green circles) for $\xi_0 = 7$ and $\beta = 10^{2.5}$. For reference, the density plot shows the absolute deviations of $R(\zeta)$ in Eq. (5.23) from unity, with the white region corresponding to $|R(\zeta) - 1| > 10^2$. The solutions shown here are those with a positive imaginary part, while their complex conjugates, ζ^* , are also valid eigenvalues of the LGEF system. *This figure is reproduced from Ref. [1].*

the AS solution for this parameter space point is unstable. However, we can also note that the two determinations of the ζ -spectrum deviate quantitatively, albeit the deviation is more pronounced for ζ 's with negative real parts. This deviation can be traced back to the approximations used by Ref. [131] leading to Eq. (5.23). In fact, we have checked explicitly that the LGEF perfectly identifies the ζ value with the largest real part that is found when solving the transcendental equation which Eq. (5.23) approximates. This makes us confident that the LGEF results are fully reliable, and in fact constitute an improvement over the results obtained from Eq. (5.23).

Next, we broaden our scope to the entire range of parameters spanned by $\xi_0 \in [5, 9]$ and $\beta \in [10^{1.5}, 10^{3.5}]$. In Fig. 5.2, we show both the real part (left panel) and imaginary part (right panel) of the Lyapunov exponent with the largest real part (henceforth called ζ_1) as a function of ξ_0 . The results using the LGEF are depicted as a blue band indicating variations in β . For comparison, we also show the results of Ref. [131] based on the roots of Eq. (5.23) in green. The left panel definitively demonstrates that, for all parameter configurations in question, $\text{Re}(\zeta_1)$ remains positive. Therefore, the AS solution appears to be unstable for our entire region of interest. Again, this conclusion follows from the LGEF, but also holds following the approach of Ref. [131]. However, the LGEF substantiates this conclusion, seeing as its results for $\text{Re}(\zeta_1)$ exceed that of Ref. [131] by an order-one factor.

Looking at the right panel of Fig. 5.2 allows for further insight into the growing perturbation mode described by ζ_1 . Its imaginary part is clearly non-vanishing for all parameter points in question. This suggests that deviations from the AS solution oscillate around the steady-state $\xi = \text{const.}$ with an amplitude that grows with time. This observation is particularly relevant as it

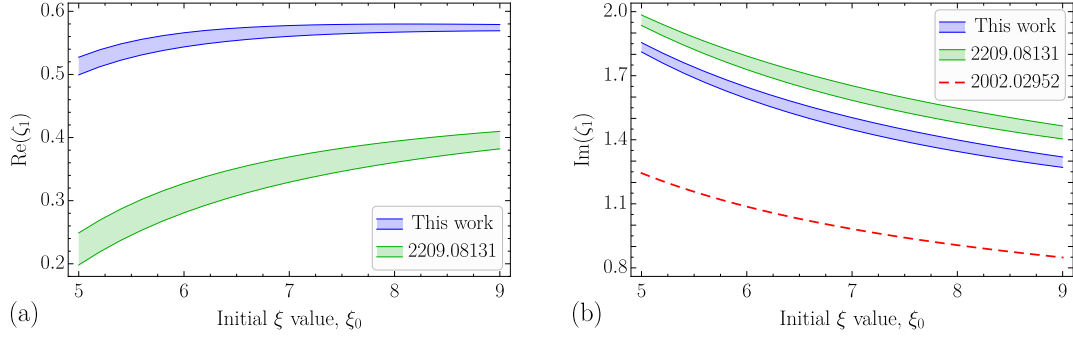


FIGURE 5.2: Results for the eigenvalue with the largest real part, ζ_1 , as a function of the instability parameter ξ_0 . Panel (a) shows the real part of ζ_1 , panel (b) its imaginary part. The bands indicate variations to these values for $\beta \in [10^{1.5}, 10^{3.5}]$. The blue band gives the results of the LGEF, and the green band that of Ref. [131] (i.e., Eq. (5.23)). The red dashed line in panel (b) shows the analytical estimate for $\text{Im}(\zeta_1)$ in Eq. (5.24) and represents the findings of Ref. [109].

This figure is reproduced from Ref. [1].

aligns with the idea that the AS solution’s instability is rooted in the retardation effects between a change in the inflaton velocity and the adjustment of gauge-field friction, $\langle \mathbf{E} \cdot \mathbf{B} \rangle$. This retarded response suggests that it is challenging to sustain a constant source of gauge-field friction, thus resulting in an oscillatory behavior of the system, precisely as suggested by $\text{Im}(\zeta_1) \neq 0$. The timescale of this retarded response for a constant value of ξ_0 was estimated in Ref. [109] to be $\Delta N_\xi \simeq \ln(\xi_0^2/2)$, which suggests an oscillatory motion of the system with a frequency

$$\omega = \text{Im}(\zeta_1) \simeq \frac{\pi}{\ln(\xi_0^2/2)}. \quad (5.24)$$

As a second point of comparison, we also show this estimate for $\text{Im}(\zeta_1)$ as a red dashed line in Fig. 5.2. Again, we find good qualitative agreement with our results for ζ_1 based on the LGEF. This observation gives further credence to the fact that our LGEF approach allows us to draw consistent conclusions about the AS solution’s instability.

5.3.2 The decay of the Anber–Sorbo solution

The results of the LGEF method elegantly demonstrate the instability of the AS solution. To expand upon these results, let us now consider this instability from a dynamical perspective. For this purpose, we study the evolution of perturbations away from the AS solution by numerically solving the evolution equations of our toy model using the GEF.

In what follows, we work with a system which, at an initial time $N = 0$, is precisely tuned to the AS solution for a constant value of the instability parameter, ξ_0 . This means, we tune all initial values of the functions e_n , g_n and b_n as in Eq. (5.13) for $\bar{\xi} = \xi_0$. We then perturb these initial conditions. In principle, such a perturbation can be arbitrary; we may perturb the instability parameter, any of the functions e_n , g_n and b_n , or a linear combination of the above. To be specific, we assume a small perturbation to the instability parameter, that is, we initialize it as $\xi(N = 0) = \xi_0 \pm \delta\xi_0$, with $\delta\xi_0/\xi_0 \ll 1$.⁴

Following our above discussion of the Lyapunov exponents for our toy model, we can expect that such a perturbation to ξ_0 dynamically grows with time, thus leading to the eventual decay of the AS solution once relative deviations from it have become of order unity. To quantify this idea,

⁴To numerically solve the GEF equations in Eq. (5.12) starting from these initial conditions, we employ the numerical strategy outlined in Sec. 4.2.

we define the survival time of the AS solution as the moment of time in e -folds when relative deviations from ξ_0 have become larger than half an order of magnitude,

$$N_{\text{AS}} \equiv \min_i N_i, \quad \frac{|\xi(N_i) - \xi_0|}{\xi_0} = 10^{-1/2}. \quad (5.25)$$

Here, N_i denotes the moments of time when the condition on the right-hand side is satisfied.

Building on our intuition from the LGEF, we can derive an analytical estimate for N_{AS} at times $N < N_{\text{AS}}$. According to the LGEF, perturbations to the AS solution are described by a linear superposition of eigenmodes, each characterized by a Lyapunov exponent ζ_n (cf. Eq. (5.21)). Therefore, the evolution of an initial perturbation to the instability parameter is then described by

$$|\xi(N) - \xi_0| = \sum_n r_n \delta\xi_0 \text{Re} \left(e^{\zeta_n N + i\varphi_n} \right). \quad (5.26)$$

Here, the real coefficients r_n , and the phases φ_n depend on how the initial perturbation $\delta\xi_0$ projects onto each eigenmode of the linear system, and on the specific moment of time which we define as $N = 0$. If there is a non-zero overlap between our choice of initial conditions and the fastest growing mode in this series (the one corresponding to ζ_1), such an eigenmode should swiftly come to dominate the evolution of ξ , such that we can approximately write

$$|\xi(N) - \xi_0| \simeq r_1 \delta\xi_0 e^{\text{Re}(\zeta_1)N}, \quad (5.27)$$

where, for simplicity, we have ignored the initial phase φ_1 . Based on this, we can then readily estimate the survival number of e -folds as

$$N_{\text{AS}} \simeq -\frac{1}{\text{Re}(\zeta_1)} \left[\ln \left| \frac{\delta\xi_0}{\xi_0} \right| + \ln \left(10^{1/2} r_1 \right) \right]. \quad (5.28)$$

To verify the estimate in Eq. (5.28), we can numerically determine the dependence of N_{AS} on the initial fine-tuning $\delta\xi_0/\xi_0$ by using the GEF to evolve $\delta\xi$ in time. In Fig. 5.3, we show the corresponding results for $\xi_0 = 7$ and $\beta = 10^{2.5}$. The analytical estimate in Eq. (5.28) is reliably reproduced by the numerical GEF results, both for positive (blue) and negative (green) initial deviations from ξ_0 . This provides a non-trivial cross-check regarding both the validity of our numerical scheme, and the results of the LGEF approach. The fact that N_{AS} decreases as a function of $\delta\xi_0/\xi_0$ also has a clear interpretation: the closer one fine-tunes the system to the exact AS solution in our toy model, the more time will pass until deviations from it have grown to order unity. However, unless the system is exactly tuned to the AS solution, it will inevitably become unstable.

As a further illustration of this dynamical instability, we show a specific example for the time-evolution of $\delta\xi/\xi_0$ in Fig. 5.4. The results are shown until the decay of the AS solution at N_{AS} for $\xi_0 = 7$ and $\beta = 10^{2.5}$ with an initial fine-tuning of $\delta\xi_0/\xi_0 = 10^{-6}$. The green solid line in this figure shows the result of the GEF, which we also use to determine the survival time $N_{\text{AS}} \approx 29$ (gray dashed lines). As a point of comparison, the blue dashed line shows the expected dynamical evolution of $\delta\xi/\xi_0$ based on the LGEF. Here, we model $\delta\xi/\xi_0$ entirely in terms of its fastest growing eigenmode, ζ_1 . The full GEF solution and the LGEF are evidently in excellent agreement: Both find that $\delta\xi$ is predominantly described by an oscillating growing mode. The agreement between these two results also highlights the degree of accuracy with which the LGEF can determine ζ_1 .

In addition to this growing mode, the GEF result in Fig. 5.4 also shows an initial stage where $\delta\xi$ decays. To explain this observation, we can note that the spectrum of Lyapunov exponents for $\xi_0 = 7$, $\beta = 10^{2.5}$ also contains decaying modes alongside the growing mode ζ_1 (cf. Fig. 5.1).

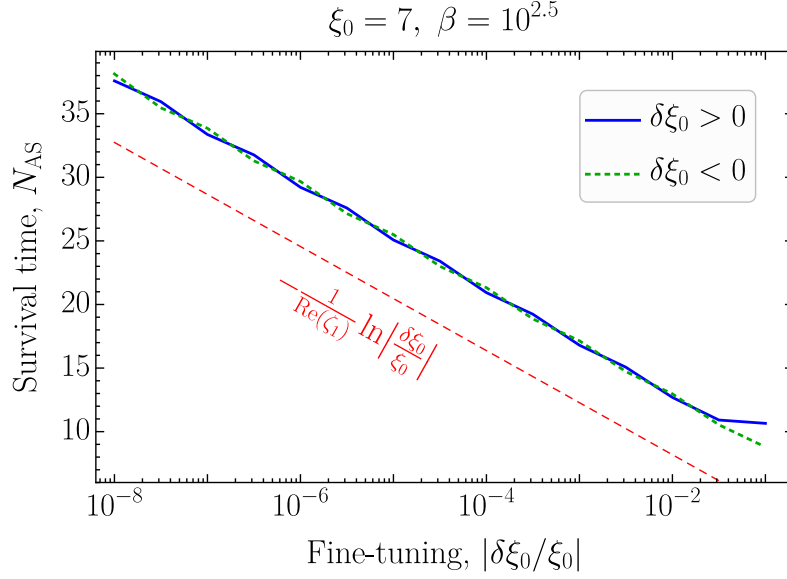


FIGURE 5.3: The survival time of the AS solution, N_{AS} , as a function of the initial fine-tuning $\delta\xi_0/\xi_0$ for $\xi_0 = 7$ and $\beta = 10^{2.5}$. The blue and green lines, respectively, correspond to positive and negative initial deviations from ξ_0 . In red, we show the analytical estimate for N_{AS} in Eq. (5.28). *This figure is reproduced from Ref. [1].*

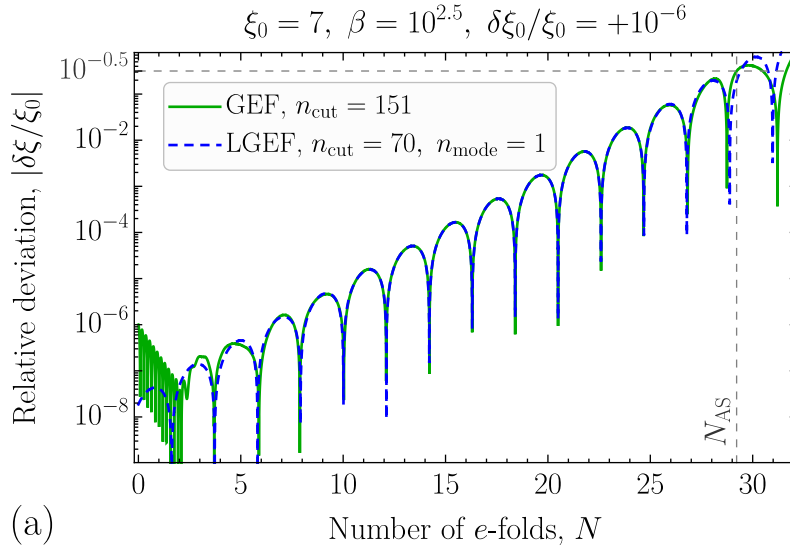


FIGURE 5.4: The time evolution of the relative deviation $\delta\xi/\xi_0$ for $\xi_0 = 7$ and $\beta = 10^{2.5}$ with an initial fine-tuning of $\delta\xi_0/\xi_0 = 10^{-6}$. The green solid line shows the GEF solution obtained for these initial conditions, while the blue dashed line shows the results of the LGEF when only including the fastest growing mode ζ_1 . The vertical gray dashed line indicates the survival time of the AS solution as defined in Eq. (5.25). *This figure is reproduced from Ref. [1].*

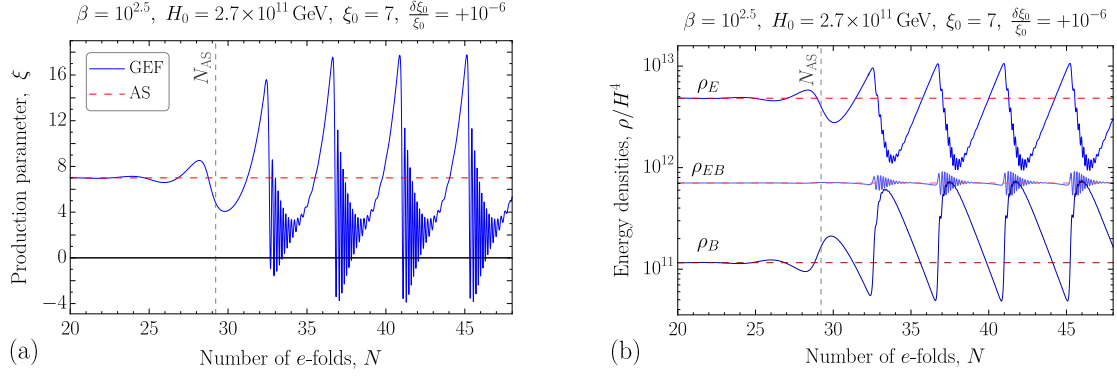


FIGURE 5.5: The time evolution of (a) the instability parameter ξ and (b) the gauge-field densities $\rho_E = \langle \mathbf{E}^2 \rangle / 2$, $\rho_B = \langle \mathbf{B}^2 \rangle / 2$, and $\rho_{EB} = |\langle \mathbf{E} \cdot \mathbf{B} \rangle| / 2$ for $\xi_0 = 7$ and $\beta = 10^{2.5}$ with an initial fine-tuning $\delta \xi_0 / \xi_0 = 10^{-6}$. The vertical dashed line shows the AS survival time in Eq. (5.25) indicating the approximate breakdown of linear perturbation theory.

This figure is reproduced from Ref. [1].

Our choice of initial conditions evidently maps the linear system onto a stable mode, which is only overcome by the unstable mode ζ_1 after a few e -folds.⁵

Finally, let us also consider how deviations from AS evolve in the nonlinear stage for $N > N_{AS}$. In this region, we may no longer rely on the linear description of our system in terms of the LGEF, and we instead rely solely on the GEF to solve the equations of motion for our axion-inflation toy model. As an illustration of this late-time behavior, we show the GEF solution for the benchmark $\xi_0 = 7$ and $\beta = 10^{2.5}$ in Fig. 5.5. In this figure, panel (a) shows the evolution of ξ , while panel (b) shows how the gauge-field energy densities, $\rho_E = \langle \mathbf{E}^2 \rangle / 2$, $\rho_B = \langle \mathbf{B}^2 \rangle / 2$, and the Chern–Pontryagin density $\rho_{EB} = |\langle \mathbf{E} \cdot \mathbf{B} \rangle| / 2$ evolve with time. The GEF results are given in blue, while the dashed red line indicates the exact AS solution. Notably, after the AS solution has decayed, the system enters into a stage characterized by a sequence of nearly perfectly periodic bursts. Using the GEF to track this behavior over multiple bursting stages, it appears to continue indefinitely.

5.3.3 Including variations in the Hubble rate

To close this discussion of the AS solution’s (in)stability, let us finally show how the above results translate to a more realistic model of axion inflation which takes into account variations in the Hubble rate via the Friedmann equation. For this case, we use the GEF from chapter 4 to numerically solve the full equations of motion for axion inflation.

When including variations of the Hubble rate, the practical initialization of the GEF in the AS regime is slightly more subtle than for a constant Hubble parameter. The issue is that the full equations of motion for axion inflation, including the Friedmann equation, do not permit a steady-state solution $\bar{\xi} = \text{const.}$ in the regime of strong backreaction. Thus, we cannot prepare a dynamical system in an exact steady-state, in particular, because gauge-field backreaction terms carry a memory of any prior changes in ξ .

Regardless, we can still initialize the GEF close to a supposed AS regime for varying H , although this construction is technically more involved. We defer a detailed description of this procedure to our original article, Ref. [1], and here only summarize its main idea. As before, we first

⁵The initial decaying phase may be accurately captured by the LGEF by also accounting for two additional modes, ζ_2 and ζ_5 (in order of decreasing real part), cf. Fig. 3, panel (b) in our original publication [1]. In particular, for $\xi_0 = 7$ and $\beta = 10^{2.5}$, we find that initial perturbations to ξ_0 map onto the stable mode ζ_5 , which explains why $\delta \xi / \xi_0$ is initially decreasing in Fig. 5.4.

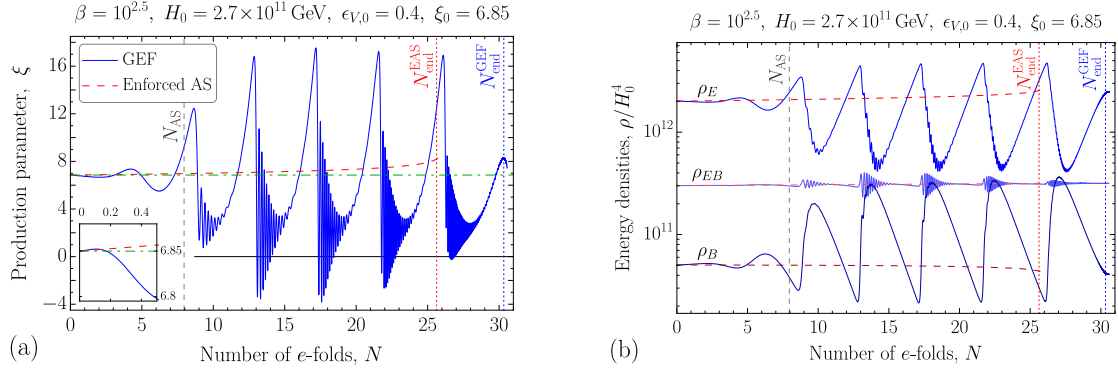


FIGURE 5.6: The time evolution of (a) the instability parameter ξ and (b) the gauge-field energy densities for a realistic model of axion inflation including the time variation of the Hubble rate. The initial conditions for this system are constructed based on an axion–vector coupling $\beta = 10^{2.5}$, an initial Hubble rate $H_0 = 2.7 \times 10^{11}$ GeV, and an initial instability parameter $\xi_0 = 6.85$. The inflaton potential gradient is adjusted such that the initial slow-roll parameter is $\epsilon_V = 0.4$ (cf. Eq. (1.30)). In blue, we show results obtained using the GEF. The red line shows the enforced AS solution (see footnote 6). The vertical gray line shows the breakdown of perturbation theory at $N = N_{AS}$ while the vertical red and blue line indicates the end of inflation for the GEF solution (blue) and the enforced AS solution (red). *This figure is reproduced from Ref. [1].*

select a target initial value for the instability parameter ξ_0 and a value for the axion–vector coupling β . For these, we may determine an initial guess for the Hubble rate, H_0 , following Eq. (5.22). This way, we ensure that the system we are studying is indeed initialized in the strong-backreaction regime. We then use ξ_0 to define the initial conditions for the GEF variables $\mathcal{X}^{(n)}$ according to Eq. (5.13). Furthermore, we use the initial guess for H_0 to determine the amplitude of the inflaton potential, V_0 in Eq. (5.2), and we adjust the potential gradient, $V_{,\phi}$, to balance any initial gauge-field friction, $\mathcal{G}^{(0)}(N = 0)$, against Hubble drag. However, the initial conditions obtained in this way do not exactly solve the equations of motion for axion inflation. Instead, one needs to initialize the system at a slightly earlier moment of time for a slightly smaller inflaton velocity, $\xi < \xi_0$, to smoothly role into a consistent set of initial conditions.

An example GEF solution that one obtains using this construction is shown in Fig. 5.6. The initial parameter configuration is $\xi_0 = 6.85$, $H_0 = 2.7 \times 10^{11}$ GeV, $\beta = 10^{2.5}$. The figure again depicts both the evolution of the instability parameter ξ in panel (a), and the gauge-field densities ρ_E , ρ_B and ρ_{EB} in panel (b), analogous to the results for constant H in Fig. 5.5. We can immediately recognize the qualitative similarity between these two figures: deviations from constant ξ (green dash-dotted line) are described by oscillations with a growing amplitude which eventually become of order unity. Once this happens, the system enters a stage of rapid recurring bursts in the inflaton velocity and the gauge-field densities. This bursting stage however only lasts for a finite duration, as inflation ultimately ends about 30 e -folds after the system has been initialized.⁶

Notably, the survival time of the AS solution for this benchmark point remains far smaller than what was achievable in the constant- H toy model of axion inflation, remaining clearly below 10 e -foldings (cf. Fig. 5.3). This observation also holds true when studying variations to the initial instability parameter ξ_0 (see Figs. 5 & 7 in Ref. [1]). We can readily attribute this to the fact that

⁶The red curve labelled “enforced Anber–Sorbo solution” (EAS) in Fig. 5.6 shows a would-be steady-state evolution in the strong-backreaction regime of axion inflation if there were no retardation between gauge-field backreaction and small changes to the instability parameter ξ . The EAS is not a true solution to the full equations of motion for axion inflation, and we construct it as a reference trajectory for the GEF in our original publication, Ref. [1] to which we refer for more details on the EAS.

the AS solution is not a true steady state if H is free to vary, thus we can never fine-tune our system to such a configuration.⁷

Overall, our results are therefore fully consistent with the claim that the AS solution is unstable. The LGEF elegantly demonstrates this instability for our toy model. This result is then further supported by the dynamical analysis of this toy model based on the GEF, and is then again consistently confirmed when we include variations to the Hubble parameter. In fact, this instability appears to be only more pronounced in the latter case. Particularly noteworthy is also that both GEF applications, once to the toy model of axion inflation and once to the more realistic axion inflation model, reproduce the same non-trivial pattern of rapid oscillations once the AS solution has decayed. This underpins the notion that our conclusions based on the toy model of axion inflation are indeed transferable to the more complete model of axion inflation which includes variations to the Hubble rate. This can be understood by noting that variations to the Hubble rate during inflation should remain small (cf. Eq. (1.19)).

5.4 More insights from the LGEF

In our above analysis, the LGEF proved to be a particularly effective tool which allowed us to definitively verify the instability of the AS solution by relying on simple linear algebra techniques. In a sense, the analysis of the Lyapunov exponents in Fig. 5.2 already demonstrates the AS instability, which our subsequent analysis of the AS solution using the full GEF validated; first for the simple toy model of axion inflation, and then in the complete model which also accounts for the drift in the Hubble rate during inflation. From this perspective, we can also see our above results as verifying that we can use the LGEF to draw consistent conclusions about the dynamics of axion inflation.

This observation raises the question if the LGEF also has other applications beyond demonstrating the instability of the AS solution. As we shall demonstrate in this section, it indeed does. In what follows, we give a brief exposition to new results based on the LGEF from our most recent article, Ref. [6], where we follow up on our original publication, Ref. [1], on which the results in Sec. 5.3 are based. Note that the complete discussion of this article is beyond the scope of this thesis, and the results we show below are only intended as a further illustration of the LGEF.⁸

Counterintuitively, although the AS stability analysis served as a motivation to develop the LGEF, and demonstrated its effectiveness, it is also the particular application of the LGEF for which it is perhaps the least suited: Like the GEF, the LGEF does not model inhomogeneous backreaction (cf. Sec. 4.3), and therefore this method should only be carefully applied to the strong-backreaction regime, where the AS solution happens to be realized. Of course, neglecting inflaton inhomogeneities was fully consistent with the question we had raised above, seeing as the AS solution was derived under the same assumption, but this does mean that our conclusions from Sec. 5.3 do not seamlessly transfer to the full axion inflation model, as inhomogeneous-backreaction effects could have stabilizing effects, which the LGEF has missed. Similarly, the rapid bursting behavior we observe after the AS solution has decayed is likely altered by inhomogeneous backreaction; based purely on Fig. 4.3, we can expect inhomogeneous backreaction to dampen this bursting behavior. In this sense, a more reasonable application of the LGEF is to use it as a diagnostic tool for studying the regime of weak backreaction, and the transition region between

⁷Notably, the bursting stage after the decay of the AS solution is also not as pronounced for larger initial values of ξ_0 . In these cases, one requires a steeper inflaton potential to realize the AS regime, and inflation ends soon after the AS solution has decayed, which prevents the onset of bursting behavior, see Ref. [1].

⁸For instance, Ref. [6] also includes an in-depth analysis of the bursting behavior appearing in Figs. 5.5 and 5.6.

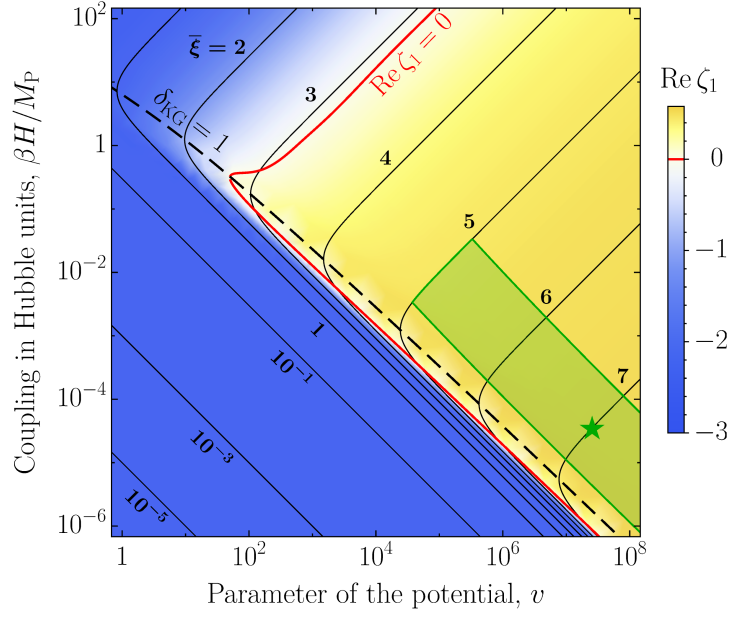


FIGURE 5.7: A map showing the stability of the steady-state solutions $\bar{\xi}(v, b)$ in Eq. (5.8) with v and b defined in Eq. (5.4). We indicate the stability of these solutions in terms of $\text{Re}(\zeta_1)$, where ζ_1 is the Lyapunov exponent for $\bar{\xi}(v, b)$ with the largest real part determined by the LGEF method. The red line indicates $\text{Re}(\zeta_1) = 0$ and black solid lines indicate contours of constant $\bar{\xi}$. The green-shaded area shows the parameter space analyzed in Sec. 5.3 with the green star indicating $\bar{\xi} = 7$, $\beta = 10^{2.5}$. The black dashed line shows $\delta_{\text{KG}} = 1$. We thank Oleksandr Sobol for adapting this figure from Ref. [6] to suit this thesis.

weak and strong backreaction, where we know our GEF description of axion inflation remains valid.

In this spirit, we follow up on the results in Sec. 5.3 by using the LGEF for a more complete stability analysis of the manifold of steady-state solutions in our axion-inflation toy model (i.e., the ones defined by Eq. (5.8)). In particular, this also includes those steady-state solutions which exist in the weak-backreaction regime, and more importantly those that exist in the transition region between weak and strong backreaction. The results of this LGEF analysis are shown in Fig. 5.7. Therein, we show the real part of ζ_1 , the Lyapunov exponent with the largest real part for $\bar{\xi}(v, b) = \text{const.}$ in Eq. (5.8). We can note two distinct regions corresponding to stable solutions, $\text{Re}(\zeta_1) < 0$ (blue), and unstable solutions $\text{Re}(\zeta_1) > 0$ (yellow), which are clearly separated by a continuous line where $\text{Re}(\zeta_1) = 0$ (red). In particular, the parameter region we studied in Sec. 5.3 is well within the unstable region (green-shaded area), while all constant- ξ solutions in the weak-backreaction regime, i.e., the ones corresponding to the conventional slow-roll attractor with $\delta_{\text{KG}} \ll 1$, are stable, exactly as we may have expected it.

We want to focus on two new features which Fig. 5.7 reveals that we could not anticipate based on our prior discussion in Secs. 5.2 and 5.3. Both are related to the fact that the conventional condition for strong backreaction, $\delta_{\text{KG}} = 1$ (dashed black), does not, in actuality, coincide with the line separating stable from unstable solutions. Indeed, we can see a region where $\delta_{\text{KG}} \gg 1$, corresponding to a regime of potentially *stable backreaction* for $\bar{\xi} \sim 1$ (upper left corner in Fig. 5.7), which our previous analysis in Sec. 5.3 did not identify. The other observation is that LGEF also identifies unstable regions which lie below the conventional criterion $\delta_{\text{KG}} = 1$, i.e., the regions where backreaction effects destabilize the weak-backreaction slow-roll attractor without gauge-field friction being more important than Hubble friction.

Let us first comment on the regime of stable backreaction. This qualitatively new dynamical regime of axion inflation has thus far not been identified by other numerical methods,⁹ and it may be seen as a region of the (v, b) parameter space in which the AS solution is stable, if we indeed define the AS solution as in Sec. 5.2.

What makes this observation especially intriguing is that in this regime of stable backreaction, inhomogeneous-backreaction effects could in fact only have a marginal effect. To understand this non-trivial point, let us reflect on the interpretation of the linear perturbations in the LGEF, Eq. (5.15), in particular, the perturbation of the instability parameter, $\delta\xi$. In the regime of stable backreaction, the LGEF suggests that the background value $\bar{\xi}$ is stable to fully homogeneous perturbations $\delta\xi(t)$, with any initial perturbations remaining small. We can then speculate that the smallness of $\delta\xi(t)$ also indicates that inhomogeneous inflaton perturbations are small as well. The basis of this speculation is to view the homogeneous perturbation, $\delta\xi(t)$, as the zero-wavenumber mode, $k = 0$ of the full spectrum of inhomogeneous perturbations with wavenumbers $k > 0$. This perspective is insightful, because we know that the spectrum of gauge-field-induced scalar perturbations is nearly scale-invariant as long as the instability parameter, $\bar{\xi}$ remains nearly constant (see, e.g., Refs. [57, 339] or Eq. 3.33). Seeing as the LGEF suggests that $\bar{\xi} = \text{const.}$ is indeed possible in the regime of stable backreaction, it is thus conceivable that a small zero-mode perturbation, $\delta\xi(t)$, also entails small inhomogeneous perturbations with $k > 0$. In particular, this may then imply that inhomogeneous-backreaction effects are indeed unimportant in the regime of stable backreaction. Of course, this line of reasoning is purely heuristic for the moment, and a more rigorous analysis of the regime of stable backreaction is needed to confirm these claims, for instance, a combined lattice and GEF investigation of the stable-backreaction regime.

Let us turn towards our second observation in Fig. 5.7, i.e., that a small region of unstable weak backreaction exists below the line $\delta_{\text{KG}} = 1$. This observation admits a more concrete, and less speculative interpretation than that of our first observation; this result entails that the LGEF criterion $\text{Re}(\zeta_1) > 0$ may prove to be a more rigorous criterion for identifying the onset of strong-backreaction effects than $\delta_{\text{KG}} = 1$. To elaborate on this, let us read Fig. 5.7 in the context of a realistic axion inflation model where we allow H to vary and where $V(\phi)$ is a realistic slow-roll inflation potential (i.e., not a simple inclined slope as in Eq. (5.2)). Consider then a scenario where we start from conventional slow-roll initial conditions on this potential (as in Sec. 4.2.2) for a fixed value of β . Assuming backreaction effects are entirely unimportant, the potential $V(\phi)$ defines a slow-roll trajectory which directly maps as a line through the v - b plane in Fig. 5.7: at each point in time the inflaton potential has a definite height and gradient which together set a unique value for the parameters b and v (see Eq. (5.4)). Such a trajectory naturally evolves from the stable regime toward the unstable regime as the inflaton generically accelerates toward larger velocities (larger values of ξ). The LGEF criterion $\text{Re}(\zeta_1) = 0$ can then be used to determine if such a slow-roll trajectory may be destabilized by strong-backreaction effects, and Fig. 5.7 suggests that this destabilization may occur even earlier than what one may expect based on δ_{KG} . Thus, the LGEF may in fact be used to determine a new and improved condition for the onset of backreaction effects. We further expand upon this point in Ref. [6], where we show that, for two realistic inflationary potentials, $\text{Re}(\zeta_1)$ can indeed be used as reliable criterion for predicting the onset of strong-backreaction effects.

Both of the above observations thus illustrate that the LGEF is a tool whose applicability extends past the mere analysis of the AS solution, and which may be the basis for further improving our understanding of the dynamics of axion inflation in the future.

⁹See Fig. 7 in Ref. [6], where we give an overview regarding the parameter regions in the v - b plane that had been covered by previous analyses.

5.5 Summary and conclusion

In this chapter, we have revisited the longstanding question regarding the stability of the Anber–Sorbo (AS) solution to axion inflation [55], which proposes that gauge-field-induced friction in the strong-backreaction regime can balance the inflaton’s potential gradient to maintain a slow-roll evolution of the inflaton field on a steep potential. To approach this question, we have adapted the gradient-expansion formalism (GEF) into a framework that can be used for a comprehensive stability analysis of the AS solution. In this, we have expanded upon the results of Ref. [131], the authors of which had previously determined the instability of the AS solution in an analytical linear stability analysis.

For the majority of this chapter, we have worked in a controlled toy model of axion inflation, assuming a constant Hubble rate H and a constant potential gradient $V_{,\phi}$, for which the AS solution is an exact steady-state solution to the equations of motion for axion inflation, and which match the assumptions used in Ref. [131]. This controlled setting allowed us to: (i) construct the linearized gradient-expansion formalism (LGEF) to determine the spectrum of Lyapunov exponents for homogeneous and time-dependent perturbations to the AS solution, and (ii) initialize the GEF directly in the AS solution to determine the growth of small time-dependent deviations away from this steady state even past the regime of validity of linear perturbation theory.

The combined application of these two tools allowed us to confirm the instability of the AS solution in the region of parameter space that was studied in Ref. [131]. We consider this conclusion to be sound given that it is consistently reached using techniques which are significantly different in their respective assumptions. While the LGEF matches the linear perturbation analysis of Ref. [131], it works under a different set of simplifying assumptions: the authors of Ref. [131] relied on finding approximate solutions to a nonlinear transcendental equation to determine the existence of an unstable eigenmode ζ_1 in the system of perturbations around the AS solution, whereas the LGEF relies on the truncation of the infinite tower of linear equations in Eq. (5.16) to arrive at the same result. Meanwhile, the direct application of the GEF gives an independent verification of our findings, as it identifies the same time-dependent growth in perturbations around the AS solution, which we could even track past the regime of validity of linear perturbation theory to show that the instability persists.

To give further credence to this conclusion, we also applied the GEF for the complete model of axion inflation in which the Hubble rate is also evolving over time. By initializing the GEF close to the supposed AS attractor solution, we could verify that variations in the Hubble rate do not allow for its stabilization. This application of the GEF also extends upon prior numerical evidence against the stability of the AS solution, which noted that a dynamical evolution of axion inflation starting from slow-roll initial conditions never reached a state of stabilized gauge-field-induced friction. While this empirical evidence left open the possibility that the AS solution’s basin of attraction is merely small, our result puts considerable tension on this idea, as the flexible choice of initial condition with the GEF permitted us to initialize our system far closer to a possible stable AS regime. It thus stands to reason that the AS solution is indeed truly unstable, at the very least when assuming large values of the instability parameter $\xi \sim 7$, as we have focussed on for the majority of this chapter.

However, the application of the LGEF to previously unexplored regions of parameter space serves to demonstrate that the instability of the AS solution may not be universal. By constructing an instability map for the toy model version of axion inflation ($H = \text{const.}$ and $V_{,\phi} = \text{const.}$), we have noted upon a thus far unidentified regime of *stable backreaction* which exists for small values of the instability parameter ξ . Meanwhile, these results requires further scrutiny, in particular as the LGEF only reveals this regime assuming that variations of the Hubble rate are unimportant.

Such a more rigorous analysis akin to the one we have performed in Sec. 5.3 is, however, beyond the scope of this thesis.

The same instability map also indicates that the LGEF can be used to determine an improved criterion to determine the onset of strong-backreaction effects when approaching the strong-backreaction regime dynamically starting from slow-roll initial conditions. This new criterion is based on the existence of an unstable eigenmode $\text{Re}(\zeta_1) > 0$ in the LGEF system of equations in Eq. (5.16). In comparison to the prevailing criterion for strong backreaction, $\delta_{\text{KG}} = 1$, which we discussed in chapter 3, the new criterion $\text{Re}(\zeta_1) > 0$ could prove to be more stringent, as the former gives an *a posteriori* statement that strong-backreaction dynamics have set in, while $\text{Re}(\zeta_1) > 0$ rather appears to be an *a priori* criterion noting that a destabilization of gauge-field friction is imminent. For example, this is suggested by the observation in Fig. 5.4 which demonstrate that the full development of an instability in the equations of motion of axion inflation requires a finite time to take effect. In this sense, the criterion $\text{Re}(\zeta_1) > 0$ indeed could prove to be an improvement over the original criterion for strong backreaction, $\delta_{\text{KG}} = 1$. We leave the pursuit of more definitive confirmation of this claim to future work beyond this thesis.

Finally, in light of our discussion of the limitations of the GEF in the previous chapter, an open question which our analysis cannot definitively answer is whether the (in)stability of the Anber–Sorbo solution persists in the presence of large inflaton inhomogeneities which can backreact on the dynamical inflation–gauge-field system. At present, this possibility would require lattice simulations as they are uniquely able to resolve the dynamical backreaction of inflaton gradients in the regime of strong gauge-field backreaction. Such lattice studies are also needed to ultimately confirm the existence of the stable-backreaction regime we have identified using the LGEF, or to refute the speculative claim we have made in Sec. 5.4 that the stable-backreaction regime also entails that inflaton inhomogeneities remain perturbatively small.

Chapter 6

Revisiting the Schwinger effect during axion inflation

This chapter is based on Ref. [3], with a partial input from Ref. [5], in particular in Sec. 6.4. The detailed contributions of the author to these articles are outlined at the beginning of this thesis.

In the previous two chapters, we have established and applied the gradient-expansion formalism (GEF) in a simple model of axion inflation involving a single, pure Abelian gauge sector that is coupled to the inflaton field. In this chapter, we now consider the application of the GEF in the context of a more complex system in which the gauge field in axion inflation is also coupled to charge carriers, like fermions or charged scalars. In this case, the enhanced background of electromagnetic fields that is created by the inflaton field during axion inflation can then directly produce charged particle–antiparticle pairs from the vacuum via the Schwinger effect.

As we have outlined in chapter 3, the Schwinger effect can have important dynamical, and ultimately phenomenological, implications for axion inflation, as charge carriers are not only produced by the gauge field, but also backreact on their production. If this process is efficient, it can inhibit gauge-field-induced friction onto the inflaton field and also impact the rate of continued charged-particle production via the Schwinger effect [127–129]. This feedback mechanism is relevant for phenomenological predictions, for instance in scenarios where the Schwinger effect acts as production mechanism for inflationary dark matter [125, 126]. Moreover, the suppression of gauge-field amplitudes affects observables linked to gauge-field production. For example, as we shall discuss in chapter 9, a reduced amplitude in gauge fields affects their ability to source gravitational waves [121],

Given these far-reaching implications of the Schwinger effect, it is necessary to have a tool such as the GEF to dynamically account for both the creation of charge carriers via the gauge-field background, and their subsequent backreaction on gauge-field production. In Ref. [128], the GEF was already adapted for this purpose by relying on a commonly adopted effective description of Schwinger pair creation (see e.g., Refs. [104, 128, 130, 132–134]). In this approach, the effect of charge carriers is modelled as a damping conductive medium. In this chapter, we revisit this effective model, we identify two of its shortcomings, and propose a refined modelling that we subsequently implement into the GEF to enable a systematic comparison with previous literature results.

At a conceptual level, the effective description we intend to modify is based on the idea that the electric and magnetic fields that are created during axion inflation also continuously create and accelerate charge carriers along the electric field lines. Effectively, this can be viewed as the Schwinger effect inducing a charge current $\mathbf{J}$, which drains energy from the amplified gauge-field background during axion inflation (cf. Eq. (3.35)). Indeed, analytical calculations of the Schwinger effect in de Sitter space support this picture, yielding a current whose magnitude depends on the strength of the electric and magnetic fields, $\mathbf{E}$ and $\mathbf{B}$, present during inflation, i.e., $|\mathbf{J}| = f(|\mathbf{E}|, |\mathbf{B}|)$ [116–121, 397–403].

In the context of axion inflation, it is typically assumed that the inflaton field produces helical gauge fields, such that the Schwinger effect proceeds via nearly (anti)collinear electric and magnetic fields. For such field configurations, analytical results suggest a charge current that is aligned with both $\mathbf{E}$ and $\mathbf{B}$, implying $|\mathbf{J}| \propto |\mathbf{E}||\mathbf{B}|$ [119–121]. This motivates that the charge current $\mathbf{J}$ can be represented in the form of Ohm’s law, i.e., suggesting the interpretation that the Schwinger effect induces a damping conductive medium of charge carriers. Because of the assumption that the electric and magnetic fields in axion inflation are collinear, two possibilities have been considered for this Ohmic representation, either writing $\mathbf{J}$ through a purely electric conductivity [104, 128], or a magnetic one [120, 121, 132]:

$$\mathbf{J} = \sigma_E \mathbf{E} \quad \text{or} \quad \mathbf{J} = \sigma_B \mathbf{B}. \quad (6.1)$$

The ambiguity in this prescription is a direct consequence of the idealized assumption that $\mathbf{E}$ and $\mathbf{B}$ during axion inflation are collinear. As part of this chapter, we propose to relax this assumption, which allows us to derive an unambiguous Ohmic representation of the induced current which depends on both an electric and a magnetic conductivity:

$$\mathbf{J} = \sigma_E \mathbf{E} + \sigma_B \mathbf{B}. \quad (6.2)$$

The second limitation that we address concerns how the resulting conductive medium affects gauge-field modes of different wavelengths. In the standard treatment, the conductivities $\sigma_{E/B}$ are assumed to be slowly varying functions of time which affect all gauge-field modes equally, independent of their wavelength [104, 128, 130]. We suggest that this is in tension with the microscopic picture of Schwinger pair creation, where this process can be considered to take place over finite distance scales [113–115]. Modelling charge carriers as a purely time-dependent conductive medium implies that even gauge modes of wavelengths far below this finite distance scale are subjected to the damping effects of charge carriers. Our second improvement is that we modify the description of the conductivities $\sigma_{E/B}$ such that Schwinger damping is only accounted for on length scales above the typical distance scale of pair creation.

The revised version of the GEF that we develop in this chapter incorporates both these improvements. It allows us to consistently describe Schwinger damping as a scale-dependent effect, while also accounting for the fact that electric and magnetic fields in axion inflation are not necessarily created in a collinear configuration.

As a direct application of this new GEF, we study the Schwinger effect in a particular benchmark model of axion inflation. In this model, the gauge field which is coupled to the inflaton is identified with the hypercharge field of the SM, such that the Schwinger mechanism leads to the direct production of SM fermions. Based on this benchmark model, we first illustrate how the dynamics of axion inflation are affected compared to the situation without Schwinger pair creation. Then, we assess how the changes we have made to the effective model of Schwinger pair creation impact phenomenologically relevant quantities like the energy density of gauge fields and the abundance of produced charge carriers. Finally, we also comment on a novel dynamical regime of backreaction which arises in the case when gauge fields are produced abundantly enough to affect the inflaton’s evolution while still being subjected to the damping effects of the Schwinger pair creation.

To build towards these results, our discussion in this chapter starts with a review of the Schwinger effect in Sec. 6.1, focussing on the aspects that are relevant for its effective modelling within axion inflation. In Sec. 6.2, we construct the effective description and introduce our modifications. Section 6.3 is where we integrate this revised effective model into the GEF, while section 6.4 presents its application to our benchmark scenario. Finally, we conclude in Sec. 6.5 with a summary and outlook.

6.1 An introduction to the Schwinger effect

Before we begin with the construction of the effective description of the Schwinger effect in axion inflation, it is useful to briefly review the basic physical mechanism underlying this nonperturbative phenomenon in quantum field theory (QFT). What we intend in this section is to motivate why the Schwinger effect can be viewed as inducing a charge current $\mathbf{J}$, as this notion is central to the effective model of Schwinger damping that we consider in the next sections of this chapter.

To build towards this idea, we first consider the Schwinger effect in Minkowski space, focussing on establishing fundamental results for the rate of pair creation in simple electromagnetic field configurations. We then extend our discussion to the Schwinger effect in de Sitter space, which is where we encounter the idea that the Schwinger effect induces a charge current $\mathbf{J}$ in the presence of strong collinear electric and magnetic fields. In particular, we present the central formula for the induced current $\mathbf{J}$ in terms of electric and magnetic field operators, $\mathbf{E}$ and $\mathbf{B}$, which explicitly enters the effective model of the Schwinger effect in axion inflation.

6.1.1 The Schwinger effect in Minkowski space

The first to identify that strong electric fields can result in the spontaneous creation of charged particles was SAUTER [110]. The idea was then further developed for quantum electrodynamics by EULER AND HEISENBERG [111]. The first to give a complete calculation for the rate of this pair-production process was SCHWINGER [112], the namesake for this effect. Modern reviews on the Schwinger effect can be found in Refs. [113–115], which we use as the basis for our subsequent discussion, and we refer to these works and the references therein for more details.

The simplest setup in which one can study the Schwinger effect is that of a matter field with charge Q and mass m , either a complex scalar or a Dirac fermion, in a constant electric field $\mathbf{E}$. Intuitively, we can understand the origin of particle production in this setup by taking the perspective that the vacuum in QFT is not empty, but populated by virtual particle–antiparticle pairs as a consequence of the Heisenberg uncertainty principle. In the presence of an electric field, these virtual pairs of charged particles are accelerated away from one another allowing them to gain sufficient energy to become real.

If we consider the electric field to be aligned with the z -axis, $\mathbf{E} = |\mathbf{E}|\mathbf{e}_z$, it classically performs a work $eQ|\mathbf{E}|\Delta z$ on a particle of charge Q over a distance Δz . Using this simple classical analogy, one can then estimate the necessary amplitude of the electric field which is needed to create a pair of particle–antiparticle pairs from the vacuum. Considering that the electric field needs to overcome the mass gap of $2m$ to create a real pair of particles, and taking the de-Broglie wavelength $\lambda \sim 1/m$ as a characteristic length, one finds that efficient particle production requires a critical electric field strength

$$|\mathbf{E}_c| \sim \frac{m^2}{e|Q|}. \quad (6.3)$$

Quantum mechanically, the pair creation process can then be understood as a tunneling process over a potential barrier $V(z) = e|Q|\mathbf{E}z$. Via the Wentzel–Kramers–Brillouin (WKB) approximation, one can determine that the tunneling rate is exponentially suppressed for electric fields which are weaker than this critical field strength.

This exponential suppression is reflected in the rate of pair production, Γ , per unit volume V in a constant electric field [111, 112] (see also Refs. [404]),

$$\frac{\Gamma}{V} = g \frac{(eQ)^2}{(2\pi)^3} |\mathbf{E}|^2 \exp\left(-\frac{\pi m^2}{e|Q\mathbf{E}|}\right) \quad (6.4)$$

Here the factor g accounts for the number of spin states of the particle, $g = 1$ for a scalar field, and $g = 2$ for a Dirac fermion. The rate can be deduced from the transition probability between in and out vacuum states before and after the electric field is turned on, the so-called vacuum-persistence probability [404];

$$\mathcal{P}_{\text{vac}} = |\langle 0_{\text{out}} | 0_{\text{in}} \rangle|^2 = |\langle 0_{\text{in}} | e^{-i\mathcal{H}t} | 0_{\text{in}} \rangle|^2 = \exp(-wVt), \quad (6.5)$$

with $\mathcal{H}$ the Hamiltonian of the theory. In the presence of an external source, this probability is generically not equal to unity. For the case of a constant, homogeneous external electric field, the factor w may be computed by means of the WKB method or by directly computing the effective action of a Dirac fermion or a complex scalar on the classical background electric field (see Refs. [113, 115, 405]). It is given by [112, 406] (see, e.g., Ref [404])

$$w = g \frac{(eQ)^2}{(2\pi)^3} |\mathbf{E}|^2 \sum_{n=1}^{\infty} \frac{1}{n^2} \exp\left(-\frac{n\pi m^2}{e|Q\mathbf{E}|}\right) \quad (6.6)$$

with the particle-creation rate Γ/V being the first term in this series [404, 407]. It should be noted that both Eqs. (6.4) and (6.6) are clearly nonperturbative results, as they both vanish in the limit $e \rightarrow 0$. This is similar to other nonperturbative effects in QFT, like the effective potential for an axion field, as we have discussed it in chapter 2. In fact, analogous to the axion effective potential, the rate of Schwinger pair creation may also be obtained via a summation over gauge-field instantons [408]. Alternatively, Eqs. (6.4) and (6.6) can also be computed using the Bogolyubov method, which we have previously discussed in Sec. 1.3 in the context of gravitational particle production during inflation.

The process of pair creation from the vacuum can be further supported by an external magnetic field, $\mathbf{B}$. Calculations of the pair-creation rate in the presence of both constant, homogeneous electric and magnetic fields are usually performed in a frame where the two fields are either collinear or anticollinear. The advantages of such a configuration is that the mechanical effect of the electric and magnetic field are distinctly separated from one another, as is most easily understood by classical considerations about the motion of a charged particle through such an electromagnetic background. When the two fields are (anti)collinear, the electric one leads to a transfer of momentum parallel to its field lines, $\mathbf{p}_{\parallel} \rightarrow \mathbf{p}_{\parallel} + q\mathbf{E}t$, while the magnetic one only affects the particles motion in the transverse direction, $\mathbf{p}_{\perp}$. In the quantum system, the electric field is then still associated with a tunneling process over a potential barrier, while the magnetic field instead implies a splitting of a field's transverse momentum, $\mathbf{p}_{\perp}$ into discrete Landau levels (see, e.g., Ref. [120, 121]). Thus, the electric field still needs to surpass the critical threshold for pair creation to be efficient, but the magnetic field increases the number density of available states which can be occupied by the newly produced particles. As a result, the overall rate of pair creation is enhanced by the magnetic field, but it features the same exponential suppression factor which is dictated by the electric field strength [406, 409];

$$\frac{\Gamma}{V} = \frac{(eQ)^2}{4\pi^2} |\mathbf{E}| |\mathbf{B}| \psi\left(\frac{\pi|\mathbf{B}|}{|\mathbf{E}|}\right) \exp\left(-\frac{\pi m^2}{e|Q\mathbf{E}|}\right). \quad (6.7)$$

Here, the function $\psi(x)$ is depends on the nature of the particle, $\psi(x) = 1/(2 \sinh x)$ for a complex scalar or $\psi(x) = \coth x$ for a Dirac fermion. This rate may again be computed using the same methods as in the absence of the magnetic field.

The Schwinger effect in Minkowski space has also been studied in various other configurations, including for alternating electric fields [410, 411] or fields at finite temperature [161, 412]. Recently,

also stochastic gauge-field configurations have been studied [413]. For our purposes, the results based on the constant collinear magnetic and electric fields are however sufficient, since this corresponds to the gauge-field configurations which we expect to be approximately realized in the case of axion inflation (see, e.g., Eq. (3.31)). However, in this case, Schwinger pair production via such a gauge-field configuration does not take place in a flat Minkowski spacetime, but rather in an inflationary spacetime which is better approximated by de Sitter space.

6.1.2 The Schwinger effect in de Sitter space

Before discussing results for the Schwinger effect in de Sitter space, let us dwell a bit further on the implications of Schwinger pair production in Minkowski space. We have seen that a constant electric field, with or without a supporting (anti)collinear magnetic field, results in a constant rate of particle creation, Eqs. (6.4) and (6.7). Considering that the electric field in this setup permeates spacetime, the creation of these charged particles is followed by their continuous acceleration. As a result, we can consider that the Schwinger effect induces a charge current $\mathbf{J}$ which is directed along the electric-field direction, and which one can estimate to be linearly growing with time (see, e.g., Refs. [114, 399, 414, 415]);

$$\mathbf{J} \simeq 2e|Q|\frac{\Gamma}{V}\frac{\mathbf{E}}{|\mathbf{E}|}t. \quad (\text{Minkowski space}) \quad (6.8)$$

Here, the rate Γ/V can either be given by Eq. (6.4) or Eq. (6.7), depending on whether a magnetic field is simultaneously present or not. In particular, it is worth reiterating that an (anti)collinear magnetic field does not affect the motion of particles along the electric field lines, thus it also does not influence the directionality of the induced current. Evidently, for $t \rightarrow \infty$, the current in Eq. (6.8) diverges, since in Minkowski space, the electric field can in principle create and accelerate particles for as long as it remains active.

Calculations of the Schwinger effect in de Sitter space are typically also expressed in terms of the analogous induced current $\mathbf{J}$. This current has been computed for a constant electric field, both for charged scalars [116, 117, 399–403], and for fermions [118, 397, 398, 402]. In the limit of a strong electric field, $e|Q\mathbf{E}| \gg H^2$, the magnitude of the induced current in this configuration is given by (see, e.g., Refs [399, 402])

$$|\mathbf{J}| = g \frac{(e|Q|)^3}{12\pi^3 H} |\mathbf{E}|^2 \exp\left(-\frac{\pi m^2}{e|Q\mathbf{E}|}\right), \quad (\text{de Sitter space, } \mathbf{E} \text{ only}) \quad (6.9)$$

with, again $g = 1$ for a scalar field, and $g = 2$ for a Dirac fermion.

The corresponding result for an additional (anti)collinear magnetic field $\mathbf{B}$ has also been computed in Refs. [119–121], with the current in the strong-field regime having a magnitude

$$|\mathbf{J}| = \frac{(e|Q|)^3}{6\pi^2 H} |\mathbf{E}||\mathbf{B}| \psi\left(\frac{\pi|\mathbf{B}|}{|\mathbf{E}|}\right) \exp\left(-\frac{\pi m^2}{e|Q\mathbf{E}|}\right), \quad (\text{de Sitter space, } \mathbf{E} \text{ and } \mathbf{B}) \quad (6.10)$$

where, again, $\psi(x) = 1/(2 \sinh x)$ for a charged scalar, or $\psi(x) = \coth x$ for a fermion.

For both field configurations, the current in Eqs. (6.9) and (6.10) is again aligned with the electric field $\mathbf{E}$. It is instructive to compare these results to the induced current in Minkowski space, Eq. (6.8) for the respective pair creation rates given by Eq. (6.4) or (6.7). The current in de Sitter space for each field configuration is indeed equivalent to the corresponding one in Minkowski space evaluated at a time $t = 1/(3H)$, implying that it is constant instead of divergent. This difference can be readily attributed to the Hubble horizon in de Sitter space; while new particles

are continuously created and accelerated by the electric field, other particles are constantly exiting the horizon, and cease to contribute to $\mathbf{J}$ (see, e.g., Ref. [117]).

These expressions for the induced current in Eqs. (6.9) and (6.10) are the foundation of the effective description of the Schwinger effect in axion inflation, as they establish the basic physical idea that the Schwinger effect induces a charge current, which can then enter Ampère’s law as in Eq. (3.35) to describe the backreaction of Schwinger pair production onto the gauge-field background in this model. It is thus important to note that they indeed permit a representation in terms of Ohm’s law, as we have hinted at in the chapter introduction. Both with and without a magnetic field, we can write the induced currents in Eqs. (6.9) and (6.10) as $\mathbf{J} = \sigma_E \mathbf{E}$. Meanwhile, in the case where we also consider an (anti)collinear magnetic field, $\mathbf{B}$, the current $\mathbf{J}$ can be equally represented as $\mathbf{J} = \sigma_B \mathbf{B}$, with σ_B being a magnetic conductivity.

Before we proceed with our discussion of axion inflation, let us also briefly comment on the induced current $\mathbf{J}$ in de Sitter space for weak fields, $q|Q\mathbf{E}| \ll H^2$. The Schwinger current in this regime has been noted to show rather counter-intuitive behavior. For light scalar fields, the Schwinger current exhibits an infrared hyperconductivity, i.e., a growing conductivity for weaker electric fields [117, 119, 400]. For fermions, the conductivity can even briefly become negative [397, 402]. In particular the negative conductivity is the subject of ongoing discussions in the literature. For instance, the authors of Ref. [416] argued that this effect is spurious and attributed it to an improper treatment of the running gauge coupling e and higher-order corrections to the gauge-field action. In Refs. [417, 418], it was recently argued that the assumption of a constant electric field in de Sitter space, which is underlying these computations, is only valid when the field is continuously replenished to counteract Hubble friction. Accounting for such a source (one of the possibilities being a rolling axion field during inflation), they confirm the results for the strong-field regime and the hyperconductivity in the weak-field regime, but they find that the fermionic Schwinger current remains positive definite.

Although these topics are highly relevant, and intrinsically fascinating, the impact of hyperconductivity on gauge-field dynamics are negligible with respect to Hubble friction [419]. For practical applications to axion inflation, where we are interested in the dynamical backreaction of Schwinger pair production, one therefore typically relies on the expressions in the strong-field regime [104, 120, 121, 128], in particular the current for (anti)collinear magnetic and electric fields in Eq. (6.10).

6.2 An effective treatment of the Schwinger effect in axion inflation

In this section, we introduce how the Schwinger effect can be incorporated into the basic model of axion inflation which we described in chapter 3. The original formulation of this effective description which we give here has been primarily developed in Ref. [127, 128, 420], and relies on the Ohmic representation of the Schwinger current in Eq. (6.10). However, as described in the introduction, we go beyond the results of Ref. [127, 128, 420] to also incorporate several improvements to their description based on our work in Ref. [3].

To differentiate between the original model of axion inflation, where charge carriers are not accounted for, and the one we present here, we refer to the former as *pure axion inflation*, highlighting that it assumes a pure gauge-field sector. In contrast, we can refer to the model with charge carriers as *fermionic axion inflation*. Although our discussion initially applies to both the case where charged scalars or fermions can couple to the gauge field, we later on focus primarily on fermions, hence the name.

6.2.1 Axion inflation with charge carriers

To describe the effect of Schwinger pair production in axion inflation, we can start from a modified version of the basic action describing axion inflation, as we have given it in Eq. (3.7). In addition to the pseudoscalar inflaton field, ϕ , and the $U(1)$ gauge field A_μ , we now want to also describe the effect of a matter field, χ , which is charged under the $U(1)$ symmetry group. Generically, this action can be written as

$$S = \int d^4x \sqrt{-g} \left[\frac{1}{2} g^{\mu\nu} \partial_\mu \phi \partial_\nu \phi - V(\phi) - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} - \frac{\beta \phi}{4M_P} F_{\mu\nu} \tilde{F}^{\mu\nu} + \mathcal{L}_\chi(\chi, A_\mu) \right]. \quad (6.11)$$

As before, we have that $V(\phi)$ is the inflaton potential, $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is the field-strength tensor with $\tilde{F}^{\mu\nu} = \varepsilon^{\mu\nu\alpha\beta} / (2\sqrt{-g}) F_{\alpha\beta}$ its dual, and β gives the strength of the axion–vector coupling. The novel term appearing in this action is the Lagrangian $\mathcal{L}_\chi$, which describes the field χ and its interaction with A_μ . For the effective description we employ, the precise form of $\mathcal{L}_\chi$ is not important. The reader may for instance imagine the Lagrangian for a charged scalar field or a Dirac fermion.

Given the close resemblance to the axion inflation model which we described in chapter 3, the next steps now proceed analogously to our discussion in Sec. 3.2. First, we derive the Euler–Lagrange equations for the gauge field and the inflaton field from Eq. (6.11). The system of equations one obtains in this way is the same as for the basic axion inflation model given in Eq. (3.8). The only modification is to the equation of motion for $F_{\mu\nu}$, which now reflects the interaction between A_μ and χ

$$\frac{1}{\sqrt{-g}} \partial_\alpha \left[\sqrt{-g} g^{\mu\alpha} g^{\nu\beta} F_{\mu\nu} \right] + \frac{\beta}{M_P} \tilde{F}^{\alpha\beta} \partial_\alpha \phi = J^\beta, \quad (6.12)$$

where J^μ is an electric four-current defined through

$$J^\mu = -\frac{\partial \mathcal{L}_{\text{ch}}}{\partial A_\mu}. \quad (6.13)$$

Furthermore, we can derive the energy–momentum tensor for this system, which reads;

$$T_{\mu\nu} = \partial_\mu \phi \partial_\nu \phi - g^{\alpha\beta} F_{\mu\alpha} F_{\nu\beta} - g_{\mu\nu} \left[\frac{1}{2} \partial_\alpha \phi \partial^\alpha \phi - V(\phi) - \frac{1}{4} F_{\alpha\beta} F^{\alpha\beta} \right] + (T_\chi)_{\mu\nu}, \quad (6.14)$$

where $(T_\chi)_{\mu\nu}$ is the energy–momentum tensor for χ which appears in addition to the contributions to $T_{\mu\nu}$ from the inflaton field and the gauge field given in Eq. (3.9).

We now assume spacetime is described by the FLRW metric in Eq. (1.23). For the electromagnetic field, this means we can rely on the same definitions for electric and magnetic field operators as in Eq. (3.10), i.e., $\mathbf{E} = -\frac{1}{a} \partial_0 \mathbf{A}$ and $\mathbf{B} = \frac{1}{a^2} \text{rot} \mathbf{A}$. Additionally, we assume that the inflaton dynamics are purely governed by the homogeneous vacuum expectation value of the inflaton field, $\varphi(t) = \langle \phi(t, \mathbf{x}) \rangle$. Furthermore, FLRW space allows to decompose the four-current J^μ in Eq. (6.13) into an electric three-current $\mathbf{J}$ and a charge density ρ_c ,

$$J^\mu = (\rho_c, \mathbf{J}/a). \quad (6.15)$$

Our working assumption is that charge carriers are produced by Schwinger pair production, such that $\mathbf{J}$ can be interpreted as the corresponding induced current, as we shall explore in more detail further below. This assumption is also important regarding the charge density ρ_c . Since Schwinger

pair creation produces particles and antiparticles in pairs, it does not result in the production of net charge. Therefore, supposing that there is no initial charge asymmetry in our system, we can safely set $\rho_c = 0$. This way, one then arrives at the following equations of motion for the inflaton field φ and the electromagnetic fields;

$$\ddot{\varphi} + 3H\dot{\varphi} + V_{,\phi}(\varphi) = \frac{\beta}{M_{\text{P}}} \langle \mathbf{E} \cdot \mathbf{B} \rangle , \quad (6.16a)$$

$$\text{div } \mathbf{E} = 0 , \quad \text{div } \mathbf{B} = 0 , \quad (6.16b)$$

$$\dot{\mathbf{E}} + 2H\mathbf{E} - \frac{1}{a} \text{rot } \mathbf{B} + \frac{\beta}{M_{\text{P}}} \dot{\varphi} \mathbf{B} + \mathbf{J} = 0 , \quad (6.16c)$$

$$\dot{\mathbf{B}} + 2H\mathbf{B} + \frac{1}{a} \text{rot } \mathbf{E} = 0 . \quad (6.16d)$$

The important modification with respect to the equations of motion for pure axion inflation is the appearance of the charge current $\mathbf{J}$ in Ampère's law, Eq. (6.16c).

The equations of motion are supplemented by the Friedmann equation, which for this system is given by

$$H^2 = \frac{1}{3M_{\text{P}}^2} \left[\frac{1}{2} \dot{\varphi}^2 + V(\varphi) + \frac{1}{2} \langle \mathbf{E}^2 + \mathbf{B}^2 \rangle + \rho_{\chi} \right] . \quad (6.17)$$

In addition to the standard contributions to the background energy density coming from the inflaton field and the electromagnetic field, the charge carriers now also contribute an energy density $\rho_{\chi} = (T_{\chi})^0_0$. Although we do not have a closed expression for ρ_{χ} , one can derive a dynamical equation describing its time evolution from covariant energy conservation, $\nabla_{\mu} T^{\mu\nu} = 0$. Taking into account the equations of motion for the electromagnetic fields and the inflaton field in Eq. (6.16), one can derive the following continuity equation for ρ_{χ} ,

$$\dot{\rho}_{\chi} + 3H(\rho_{\chi} + p_{\chi}) = \langle \mathbf{J} \cdot \mathbf{E} \rangle , \quad (6.18)$$

where $p_{\chi} = -1/3(T_{\chi})^i_i$ is the pressure associated with the charge carriers. The left-hand side can be interpreted as describing the dilution of ρ_{χ} due to cosmic expansion, while on the right-hand side we interpret the current $\mathbf{J}$ as sourcing charge carriers. Just as for ρ_{χ} , we do not have a closed expression for the pressure p_{χ} , however, in practice, one can simply consider a definite equation of state, e.g., for a massless field, $p_{\chi} = \rho_{\chi}/3$.

In total, we then have a complete dynamical system of equations which describe the evolution of axion inflation in the presence of a charged current $\mathbf{J}$. In principle, this is just a generalization of the pure axion inflation model in chapter 3, which we retrieve in the particular case of absent charge carriers, $\mathbf{J} = 0$ and $\rho_{\chi} = 0$. What we should then consider in more detail is an appropriate definition of the induced current $\mathbf{J}$ to obtain a closed set of differential equations.

6.2.2 The induced current

Our ultimate aim is to identify the charge current $\mathbf{J}$ with the Schwinger-induced current in Eq. (6.10). To start out, this requires us to reflect on the physical process of pair creation in axion inflation to motivate a suitable Ansatz for $\mathbf{J}$.

We have seen in the previous chapters that the production of electric and magnetic fields in axion inflation results in large-scale classical electric and magnetic fields. These fields have a random orientation in each Hubble patch, which average to a statistically homogeneous and isotropic ensemble over the entire Universe with their typical magnitudes given by $\sqrt{\langle \mathbf{E}^2 \rangle}$ and $\sqrt{\langle \mathbf{B}^2 \rangle}$. Schwinger pair production now occurs locally within a single Hubble patch depending

on the strength of the local classical fields $\mathbf{E}(t, \mathbf{x})$ and $\mathbf{B}(t, \mathbf{x})$. The spontaneous creation of particle–antiparticle pairs in each patch then results in a local charge current $\mathbf{J}(t, \mathbf{x})$, which can be computed using the expression for the Schwinger-induced current in Eq. (6.10). The orientation of this current then depends on the local electric and magnetic fields, but one can assume that its magnitude varies only mildly between local Hubble patches, and can be characterized in terms of the mean electric and magnetic field strengths.

This intuitive picture motivates the Ohmic description of the induced current $\mathbf{J}(t, \mathbf{x})$, which we can generally write as

$$\mathbf{J}(t, \mathbf{x}) = \sigma_E(t) \mathbf{E}(t, \mathbf{x}) + \sigma_B(t) \mathbf{B}(t, \mathbf{x}), \quad (6.19)$$

with σ_E and σ_B given as purely time-dependent conductivities depending on expectation values $\langle \mathbf{E}^2 \rangle$, $\langle \mathbf{B}^2 \rangle$ and $\langle \mathbf{E} \cdot \mathbf{B} \rangle$. The physical interpretation of Eq. (6.19) is that the charge carriers produced through Schwinger pair creation act as a global conductive medium, which locally affects the electric and magnetic fields in individual Hubble patches.

To obtain expressions for the conductivities σ_E and σ_B , one matches Eq. (6.19) to the Schwinger-induced current in Eq. (6.10). Specifically, we have the two distinct representations

$$\mathbf{J} = \sigma_E \mathbf{E} \quad \text{with} \quad \sigma_E = \frac{(e|Q|)^3}{6\pi^2 H} |\mathbf{B}| \psi\left(\frac{\pi|\mathbf{B}|}{|\mathbf{E}|}\right) e^{-\frac{\pi m^2}{e|Q||\mathbf{E}|}}, \quad (6.20a)$$

or

$$\mathbf{J} = \sigma_B \mathbf{B} \quad \text{with} \quad \sigma_B = \frac{(e|Q|)^3}{6\pi^2 H} \text{sign}(\mathbf{E} \cdot \mathbf{B}) |\mathbf{E}| \psi\left(\frac{\pi|\mathbf{B}|}{|\mathbf{E}|}\right) e^{-\frac{\pi m^2}{e|Q||\mathbf{E}|}}. \quad (6.20b)$$

Both these expressions have been considered in the literature on axion inflation, some authors working in the *electric picture*, Eq. (6.20a), (for instance Refs. [104, 128]) and others in the *magnetic picture*, Eq. (6.20b), (for instance Refs. [120, 121]). The problem with this ambiguous description is that electric and magnetic conductivities do not affect the dynamics of axion inflation in the same way. This can for example be seen by noting the different roles they play in Ampère’s law, Eq. (6.16c):

$$\dot{\mathbf{E}} + (2H + \sigma_E) \mathbf{E} - \frac{1}{a} \text{rot } \mathbf{B} + \frac{\beta}{M_P} \dot{\varphi} \mathbf{B} = 0, \quad (\text{electric picture}) \quad (6.21a)$$

$$\dot{\mathbf{E}} + 2H \mathbf{E} - \frac{1}{a} \text{rot } \mathbf{B} + \left(\frac{\beta}{M_P} \dot{\varphi} + \sigma_B \right) \mathbf{B} = 0. \quad (\text{magnetic picture}) \quad (6.21b)$$

In the electric picture, the conductivity σ_E acts as an additional source of Hubble friction, while in the magnetic picture, σ_B directly inhibits the generation of electromagnetic fields through the rolling inflaton field (generically, σ_B takes an opposite sign to $\dot{\varphi}$). Although both qualitatively result in a reduced amplitude of electromagnetic fields, the two effects are not physically equivalent.

This ambiguity merely arises because of the idealized assumption of perfectly collinear electric and magnetic fields used to derive Eq. (6.10). Numerical simulations of axion inflation in the presence of an electrical conductivity indeed demonstrate that collinearity only holds approximately, and is never truly achieved in the comoving frame [128]. We therefore propose to relax the assumption of collinearity, and to derive an expression for the Schwinger current which is valid also for non-collinear fields. As a result, the Schwinger-induced current naturally decomposes into a component aligned with the electric field, and one which aligns with the magnetic field. This allows one to represent the current $\mathbf{J}$ in a general Ohmic form as in Eq. (6.19), including both

types of conductivities. The resulting description of the Schwinger effect can be referred to as the *mixed picture* in contrast to the purely electric or magnetic pictures in Eq. (6.20).

Derivation of the mixed picture

To derive an expression for the Schwinger-induced current that is valid for non-collinear fields, we apply a simple procedure: We perform a Lorentz boost to a frame where $\mathbf{E}'$ and $\mathbf{B}'$ are collinear; there, we evaluate the expression for the induced current in Eq. (6.10); then we boost back to the comoving frame. This is a standard approach in Minkowski space to derive the rate of Schwinger pair creation for an arbitrary configuration of homogeneous and constant electromagnetic fields (see Refs. [113, 114]). It has been previously used for axion inflation in Ref. [129], a perturbative study of Schwinger pair creation. A similar method had been considered in Ref. [421], although this splitting was intended to separately consider the production of charge carriers on the lowest Landau level (i.e., the ones not influenced by the magnetic field), and those on higher Landau levels. In contrast, we derive expressions for the Ohmic conductivities in the mixed picture, which are also valid beyond perturbation theory.

We begin our computation by boosting to the collinear frame. Under a Lorentz boost with velocity $\mathbf{v}$, electromagnetic fields transform according to (see, e.g., the textbook [422])

$$\mathbf{E}' = \gamma(\mathbf{E} + \mathbf{v} \times \mathbf{B}) - \frac{\gamma^2}{1 + \gamma} \mathbf{v}(\mathbf{v} \cdot \mathbf{E}), \quad \mathbf{B}' = \gamma(\mathbf{B} - \mathbf{v} \times \mathbf{E}) - \frac{\gamma^2}{1 + \gamma} \mathbf{v}(\mathbf{v} \cdot \mathbf{B}), \quad (6.22)$$

where $\gamma = (1 - v^2)^{-1/2}$ is the corresponding Lorentz factor. To ensure that the fields in the boosted frame are collinear, we require $\mathbf{E}' \times \mathbf{B}' = 0$. This leaves infinitely many possibilities, but we want to consider the frame in which $\mathbf{E}$ and $\mathbf{B}$ lie in the same plane as $\mathbf{E}'$ and $\mathbf{B}'$. This uniquely defines $\mathbf{v}$ as the boost velocity which is perpendicular to $\mathbf{E}$ and $\mathbf{B}$. The velocity can be determined algebraically, and is given by

$$\mathbf{v} = \kappa \mathbf{E} \times \mathbf{B}, \quad \kappa = \frac{2}{\mathbf{E}^2 + \mathbf{B}^2 + \sqrt{(\mathbf{E}^2 - \mathbf{B}^2)^2 + 4(\mathbf{E} \cdot \mathbf{B})^2}}. \quad (6.23)$$

The corresponding Lorentz factor is given by

$$\gamma = \frac{1}{\sqrt{1 - v^2}} = \frac{1}{\sqrt{2}} \left[1 + \frac{\mathbf{E}^2 + \mathbf{B}^2}{\sqrt{(\mathbf{E}^2 - \mathbf{B}^2)^2 + 4(\mathbf{E} \cdot \mathbf{B})^2}} \right]^{1/2}. \quad (6.24)$$

The magnitudes of the electric and magnetic fields in the collinear frame, $\mathbf{E}'$ and $\mathbf{B}'$, are related to those in the comoving frame, $\mathbf{E}$ and $\mathbf{B}$, through the Lorentz invariants of the field-strength tensor

$$\frac{1}{2} F_{\mu\nu} F^{\mu\nu} = \mathbf{B}^2 - \mathbf{E}^2 = B'^2 - E'^2, \quad (6.25a)$$

$$-\frac{1}{4} F_{\mu\nu} \tilde{F}^{\mu\nu} = \mathbf{E} \cdot \mathbf{B} = E' B'. \quad (6.25b)$$

Here, we have defined E' as the magnitude of $\mathbf{E}'$, which is strictly positive, and B' is the projection of the magnetic field on the electric-field direction, so it can be negative for anticollinear fields. E'

and B' can then be determined entirely in terms of $\mathbf{E}$ and $\mathbf{B}$;

$$E' = \frac{1}{\sqrt{2}} \left[\mathbf{E}^2 - \mathbf{B}^2 + \sqrt{(\mathbf{E}^2 - \mathbf{B}^2)^2 + 4(\mathbf{E} \cdot \mathbf{B})^2} \right]^{1/2}, \quad (6.26a)$$

$$B' = \text{sign}(\mathbf{E} \cdot \mathbf{B}) \frac{1}{\sqrt{2}} \left[\mathbf{B}^2 - \mathbf{E}^2 + \sqrt{(\mathbf{E}^2 - \mathbf{B}^2)^2 + 4(\mathbf{E} \cdot \mathbf{B})^2} \right]^{1/2}. \quad (6.26b)$$

We now aim at finding a current $\mathbf{J} = \sigma_E \mathbf{E} + \sigma_B \mathbf{B}$ in the comoving frame. This current is orthogonal to the velocity in Eq. (6.23), so a boost along this direction leaves it invariant, $\mathbf{J} = \mathbf{J}'$. In the collinear frame, $\mathbf{J}'$ can be directly taken from Eq. (6.10), as this expression was derived assuming collinear fields:

$$|\mathbf{J}'| = \frac{(e|Q|)^3}{6\pi^2 H} E' |B'| \psi\left(\frac{\pi |B'|}{E'}\right) e^{-\frac{\pi m^2}{e|Q|E'}}. \quad (6.27)$$

This current is collinear with the electric field $\mathbf{E}'$, so we can write down the following relation

$$\begin{aligned} \mathbf{J} = |\mathbf{J}'| \frac{\mathbf{E}'}{E'} &= \frac{|\mathbf{J}'| \gamma}{E'} [\mathbf{E} + \kappa \mathbf{B} \times (\mathbf{B} \times \mathbf{E})] \\ &= \frac{|\mathbf{J}'| \gamma}{E'} [(1 - \kappa \mathbf{B}^2) \mathbf{E} + \kappa (\mathbf{E} \cdot \mathbf{B}) \mathbf{B}] = \sigma_E \mathbf{E} + \sigma_B \mathbf{B}, \end{aligned} \quad (6.28)$$

with

$$\sigma_E = \frac{|\mathbf{J}'| \gamma}{E'} (1 - \kappa \mathbf{B}^2) = \frac{|\mathbf{J}'| E'}{\gamma} \frac{1}{\sqrt{(\mathbf{E}^2 - \mathbf{B}^2)^2 + 4(\mathbf{E} \cdot \mathbf{B})^2}}, \quad (6.29a)$$

$$\sigma_B = \frac{|\mathbf{J}'| \gamma}{E'} \kappa (\mathbf{E} \cdot \mathbf{B}) = \frac{|\mathbf{J}'|}{E' \gamma} \frac{(\mathbf{E} \cdot \mathbf{B})}{\sqrt{(\mathbf{E}^2 - \mathbf{B}^2)^2 + 4(\mathbf{E} \cdot \mathbf{B})^2}}, \quad (6.29b)$$

where we used Eqs. (6.23), (6.24), and (6.26a) in the second line of each respective equation. These conductivities can finally be expressed entirely through $|\mathbf{E}|$, $|\mathbf{B}|$, and $(\mathbf{E} \cdot \mathbf{B})$ using Eq. (6.24), Eqs. (6.26), and Eq. (6.27) to evaluate γ , E' , B' and $|\mathbf{J}'|$. By replacing these by the appropriate mean-field expressions, $\sqrt{\langle \mathbf{E}^2 \rangle}$, $\sqrt{\langle \mathbf{B}^2 \rangle}$ and $\langle \mathbf{E} \cdot \mathbf{B} \rangle$, we can indeed express the current $\mathbf{J}$ in the mixed picture as given in Eq. (6.19).

Taking the results of this section, we can then return to the differential equations describing the dynamical evolution of axion inflation, Eqs. (6.16) to (6.18). With the Ohmic expression for $\mathbf{J}$ in Eq. (6.19), Ampère's law, Eq. (6.16c), becomes

$$\dot{\mathbf{E}} + (2H + \sigma_E) \mathbf{E} - \frac{1}{a} \text{rot } \mathbf{B} + \left(\frac{\beta}{M_P} \dot{\varphi} + \sigma_B \right) \mathbf{B} = 0. \quad (6.30)$$

Furthermore, the continuity equation for ρ_χ in Eq. (6.18) can now be written as

$$\dot{\rho}_\chi + 3H(\rho_\chi + p_\chi) = \sigma_E \langle \mathbf{E}^2 \rangle + \sigma_B \langle \mathbf{E} \cdot \mathbf{B} \rangle. \quad (6.31)$$

The conductivities σ_E and σ_B appearing in these equations are given entirely as functions of $\sqrt{\langle \mathbf{E}^2 \rangle}$, $\sqrt{\langle \mathbf{B}^2 \rangle}$ and $\langle \mathbf{E} \cdot \mathbf{B} \rangle$. We therefore achieve a fully closed set of differential equations which also dynamically incorporate the effect of Schwinger pair creation in axion inflation.

6.2.3 Damping in a conductive medium

We have already developed a qualitative understanding on how Schwinger pair creation affects the generation of electric and magnetic fields, noting that the electric conductivity σ_E effectively adds to Hubble friction (see Eq. (6.21a)), while the magnetic conductivity directly inhibits gauge-field generation through $\dot{\varphi}$ (see Eq. (6.21b)). We now follow up on this notion by studying how the conductive medium affects the evolution of individual gauge-field modes.

To this end, consider the quantized gauge field $\mathbf{A}$ as we have defined it in Eq. (3.15);

$$\mathbf{A}(t, \mathbf{x}) = \int \frac{d^3\mathbf{k}}{(2\pi)^{3/2}} \sum_{\lambda=\pm} \left[\epsilon^\lambda(\mathbf{k}) \hat{a}_{\mathbf{k},\lambda} A_\lambda(t, k) e^{i\mathbf{k}\cdot\mathbf{x}} + \epsilon^{*\lambda}(\mathbf{k}) \hat{a}_{\mathbf{k},\lambda}^\dagger A_\lambda^*(t, k) e^{-i\mathbf{k}\cdot\mathbf{x}} \right], \quad (6.32)$$

where the mode functions $A_\lambda(t, k)$ are again described in a circular polarization basis given by $\epsilon^\lambda(\mathbf{k})$, which obey Eq. (3.16), and the creation (annihilation) operators $\hat{a}_{\mathbf{k},\lambda}^\dagger$ ($\hat{a}_{\mathbf{k},\lambda}$) are defined to be canonically commuting. In the presence of conductivities, the mode equation for $A_\lambda(t, k)$ can be derived from Eq. (6.30), and is

$$\ddot{A}_\lambda(t, k) + (H + \sigma_E) \dot{A}_\lambda(t, k) + \left[\left(\frac{k}{a} \right)^2 - \lambda \frac{k}{a} \left(\frac{\beta}{M_P} \dot{\varphi} + \sigma_B \right) \right] A_\lambda(t, k) = 0. \quad (6.33)$$

In the previous section, we have seen that σ_E and σ_B are themselves depending on the gauge-field expectation values $\langle \mathbf{E}^2 \rangle$, $\langle \mathbf{B}^2 \rangle$ and $\langle \mathbf{E} \cdot \mathbf{B} \rangle$, and are therefore functions of $A_\lambda(t, k)$. This makes the mode equation in Eq. (6.33) highly nonlinear and enormously complicates the study of the mode functions $A_\lambda(t, k)$. The common approach in the literature (for instance Refs. [104, 128]) is therefore to treat σ_E and σ_B as given, slowly varying, external functions of time t . While this assumption greatly simplifies the analytical study of Schwinger damping, it has a definitive downside in that the conductivities $\sigma_E(t)$ and $\sigma_B(t)$ then affect all gauge-field modes $A_\lambda(t, k)$ indiscriminately of their precise wavenumber k . This conflicts with physical intuition, as it entails that gauge-field modes of any arbitrary wavenumber k can feel the presence of the conductive medium at all times. This leads to the unintuitive conclusion that even gauge-field modes whose wavelengths are much shorter than the typical separation of charge carriers in the conducting medium are affected by it. In particular, this effect even extends to modes which are effectively oscillating on scales far below the Hubble horizon, $k \gg aH$.

To avoid the counter-intuitive consequences of this effective modelling, while still maintaining a relative degree of mathematical simplicity, we propose a mild modification to it, in that we allow for an additional scale dependence of the conductivities $\sigma_E(t)$ and $\sigma_B(t)$, replacing them by

$$\sigma_{E/B}(t, k) \equiv \sigma_{E/B}(t) \Theta(t, k). \quad (6.34)$$

The window function $\Theta(t, k)$ is introduced by hand to capture the scale dependence of Schwinger damping as the system evolves with time. This is of course an *ad hoc* solution, which ultimately should be improved upon by considering a first-principles approach to the problem, for instance studying Schwinger pair creation in a constant background gauge field in the presence of high-frequency modes. However, given the current state in the literature, where Schwinger pair creation during axion inflation is treated predominantly based on an effective description, this simple modification represents a straightforward, and easy-to-implement improvement over existing approaches.

Introducing the window function Θ , the mode equation in Eq. (6.33) is then given by

$$\ddot{A}_\lambda(t, k) + (H + \sigma_E \Theta) \dot{A}_\lambda(t, k) + \left[\left(\frac{k}{a} \right)^2 - 2\lambda H \frac{k}{a} \left(\xi + \frac{\sigma_B \Theta}{2H} \right) \right] A_\lambda(t, k) = 0. \quad (6.35)$$

For later convenience, we have rewritten $\dot{\varphi}$ in terms of the instability parameter ξ , which we defined for pure axion inflation in Eq. (3.20). We now proceed to analyze this mode equation as we have done in the case of regular axion inflation in chapter 3, that is, we identify suitable initial conditions for the mode $A_\lambda(t, k)$ and consider how particle production via the tachyonic instability is affected by the presence of the conductivities. To do this, we keep the window function $\Theta(t, k)$ arbitrary for the moment and only later on discuss particular choices one can make for it.

Before we start to analyze the full mode equation, it is instructive to only consider the magnetic conductivity in isolation, as its effect is particularly easy to understand. Thus, for the moment, we set $\sigma_E = 0$ in our minds. In this case, the mode equation in Eq. (6.35) is analogous to the one in pure axion inflation, as we gave it in Eq. (3.18), the only difference being the appearance of the magnetic conductivity, which acts as an additional contribution to ξ . This motivates the definition of an effective instability parameter

$$\xi_{\text{eff}} \equiv \xi + \frac{\sigma_B}{2H} \Theta(t, k). \quad (6.36)$$

In the absence of an electric conductivity, the gauge modes $A_\lambda(t, k)$ are subjected to a similar tachyonic enhancement as in pure axion inflation, the difference now being that this only occurs for modes with $k \lesssim 2|\xi_{\text{eff}}|aH$. Since σ_B is generically of opposite sign to ξ , we have that $|\xi_{\text{eff}}| < |\xi|$.¹ So, in the presence of σ_B , fewer modes $A_\lambda(t, k)$ are exponentially amplified at a given moment of time. In this way, the magnetic conductivity directly results in reduced gauge-field production.

The additional impact of the electric conductivity is also easy to understand; each mode $A_\lambda(t, k)$ accumulates friction over time, the effect of which is encompassed by the factor

$$\Delta_k(t) \equiv \exp \left(- \int_{-\infty}^t \sigma_E(t') \Theta(t', k) dt' \right). \quad (6.37)$$

To further analyze the mode equation in Eq. (6.35) in the presence of both conductivities, it is useful to introduce the rescaled mode functions,

$$A_\lambda(t, k) = \sqrt{\Delta_k(t)} f_\lambda(t, k). \quad (6.38)$$

In conformal time, η , the modes $f_\lambda(\eta, k)$ then evolve according to

$$f_\lambda''(\eta, k) + \left\{ k^2 - 2\lambda k \xi_{\text{eff}} a H - \frac{a^2}{2} \left[(\sigma_E \Theta)' + \frac{1}{2} (\sigma_E \Theta)^2 + H \sigma_E \Theta \right] \right\} f_\lambda(\eta, k) = 0, \quad (6.39)$$

which can be further simplified by assuming that H and $\sigma_E \Theta(t, k)$ varies only slowly during inflation, such that we get

$$f_\lambda''(\eta, k) + \left[k^2 - 2\lambda k \xi_{\text{eff}} a H - \frac{s_E + s_E^2}{k^2} a^2 H^2 \right] f_\lambda(\eta, k) = 0, \quad (6.40)$$

¹This can be qualitatively seen in Eqs. (3.29) and (6.29b), which imply $\sigma_B \propto \mathbf{E} \cdot \mathbf{B} \propto -\text{sgn}(\xi)$.

where we defined the useful variable

$$s_E = \frac{\sigma_E}{2H} \Theta(t, k). \quad (6.41)$$

From Eq. (6.40), we can then see that, in the presence of an electric conductivity, it is the modes $f_\lambda(t, k)$ for which we can define Bunch–Davies initial conditions,

$$f_\lambda(\eta, k) = \frac{1}{\sqrt{2k}} e^{-i\eta k}, \quad k \gg aH, \quad (6.42)$$

while the physical gauge-field modes deep inside the horizon are given by

$$A_\lambda(\eta, k) = \sqrt{\frac{\Delta_k(\eta)}{2k}} e^{-i\eta k}, \quad k \gg aH. \quad (6.43)$$

In addition, Eq. (6.40) also highlights how the electric conductivity affects the tachyonic instability. It is still present, but now the critical scale below which a mode $f_\lambda(t, k)$ is enhanced is given by

$$\frac{k_{\text{crit}}}{aH} = |\xi_{\text{eff}}| + \sqrt{\xi_{\text{eff}}^2 + s_E^2 + s_E}, \quad (6.44)$$

The instability is still helical and only affects modes with $\lambda = \text{sgn } \xi_{\text{eff}}$. Interestingly, we have that $k_{\text{crit}} \geq 2|\xi_{\text{eff}}|aH$, which means that *more* not fewer modes $f_\lambda(t, k)$ are tachyonically enhanced in the presence of an electric conductivity. However, overall gauge-field production is still drastically reduced, as the true mode functions $A_\lambda(t, k)$ are damped by the accumulating friction given by $\Delta_k(t)$.

Choices for the window function

Having developed the general formalism as well as an improved understanding of the role played by σ_E and σ_B , we can now consider two particular choices for $\Theta(t, k)$.

The first one is $\Theta(t, k) = 1$, which corresponds to the choice made in the literature prior to our work, i.e., neglecting scale-dependent effects (e.g., in Ref. [128]). From Eq. (6.37), we see that this implies

$$\Delta_k(t) \rightarrow \Delta(t) = \exp \left(- \int_{-\infty}^t \sigma_E(t') dt' \right), \quad (6.45)$$

i.e., the factor $\Delta_k(t)$ is a global function of time and independent of k . Evidently, this entails that all modes A_λ will feel the presence of the conductivity induced by the charge carriers at all times, including those that are deep inside the horizon, see Eq. (6.43).

Let us then consider how we can make a more refined choice for $\Theta(t, k)$ by pondering on the physical nature of Schwinger pair production. As we have described in Sec. 6.1, the basic mechanism behind particle creation in a strong electric field is that the field accelerates a virtual particle–antiparticle pair until it has imparted sufficient energy onto it for the pair to become real. This process cannot happen instantaneously, so there is a characteristic minimal distance δx_S by which real charge carriers are separated in a conducting medium created by the Schwinger effect. The implication for the gauge field is that modes with wavelengths much shorter than δx_S should not feel the presence of these charge carriers, and should therefore not be subjected to any damping effect. This distance scale δx_S should thus be reflected in the window function $\Theta(t, k)$,

Let us therefore estimate the scale δx_S . To simplify this, we consider the collinear frame and assume a strong field, $m^2 \ll |eQE'|$. As previously discussed, in this frame, the electric field and

the magnetic field affect the motion of an accelerated particle in two distinct ways; the electric field accelerates the particle along the electric field lines, while the magnetic field affects the particles motion in the plane transverse to this direction. To prevent annihilation, the electric field must then separate the created particles and antiparticles in the longitudinal direction by a distance greater than the corresponding de Broglie wavelength, i.e.,

$$\delta x' > \lambda_{\text{dB}} = \frac{2\pi}{p} = \frac{2\pi}{|eQE'|t'} \simeq \frac{2\pi}{|eQE'|\delta x'}, \quad (6.46)$$

where in the last step we took into account that the created particles are ultrarelativistic.

From this, we can deduce that Schwinger pair production in the collinear frame occurs on a characteristic physical spatial and temporal scale given by

$$t'_S \sim \delta x'_S \sim \frac{1}{\sqrt{|eQE'|}}. \quad (6.47)$$

Now, we can boost back to the comoving reference frame and take into account that $\delta x'_S$ is oriented along the direction of the electric field $\mathbf{E}'$, i.e., perpendicular to the boost velocity. Then, the characteristic spatial scale in the comoving frame remains the same as in the collinear frame,

$$\delta x_S = \delta x'_S \sim \frac{1}{\sqrt{|eQE'|}}. \quad (6.48)$$

One should also consider this scale in the context of the Hubble radius. We can see that for weak fields, $|eQE'| \ll H^2$ the real particle–antiparticle pairs are seemingly created at separations larger than the Hubble radius at the given moment of time, $\delta x_S \gg H^{-1}$. This collides with our understanding of causality, so it is natural to assume that no particle creation occurs to begin with. Consequently, there should also be no damping effect due to charge carriers in this case. This is a second scale-dependent effect which we should account for in our definition of the window function $\Theta(t, k)$. Note that this reasoning also aligns with the basic assumptions in the derivation of the induced current in Eq. (6.10), which is only valid for strong fields $|eQE'| \gg H^2$.

In total, our choice for $\Theta(t, k)$ should therefore reflect that gauge modes are not damped if their wavelength is shorter than δx_S or if $\delta x_S \gg H^{-1}$. The simplest way to take this into account is to model $\Theta(t, k)$ in terms of two step functions,

$$\Theta(t, k) = \theta(k_S(t) - k) \theta(k_S(t) - a(t)H(t)), \quad (6.49)$$

where the threshold momentum $k_S(t)$ corresponds to the Schwinger pair-creation scale in Eq. (6.47). For definiteness, we define $k_S(t)$ as

$$k_S(t) \equiv a(t) \sqrt{|eQ\mathbf{E}'(t)|} = a(t) \frac{|eQ|}{2^{1/4}} \left[\mathbf{E}^2 - \mathbf{B}^2 + \sqrt{(\mathbf{E}^2 - \mathbf{B}^2)^2 + 4(\mathbf{E} \cdot \mathbf{B})^2} \right]^{1/4}, \quad (6.50)$$

where we used Eq. (6.26a) for the electric field in the collinear frame.

Let us consider the implications of this new window function $\Theta(t, k)$ in Eq. (6.49) for the evolution of a gauge-field mode $A_\lambda(t, k)$. The time evolution of a given gauge-field mode $A_\lambda(t, k)$ now separates into two stages. In the first stage, when $k > k_S(t)$, it does not feel any damping effects, and its evolution is described by Eq. (6.35) with both electric and magnetic conductivities set to zero. In conformal time, this mode equation for $k > k_S(t)$ takes the form

$$A''_\lambda(\eta, k) + (k^2 - 2\lambda\xi kaH)A_\lambda(\eta, k) = 0. \quad (6.51)$$

Deep inside the horizon, the mode $A_\lambda(t, k)$ is then the one which is in a vacuum state given by the Bunch–Davies initial condition,

$$A_\lambda(\eta, k) = \frac{1}{\sqrt{2k}} e^{-ik\eta}. \quad (6.52)$$

This means we have no unphysical damping of vacuum modes.

The damping effects due to the Schwinger conductivities remain absent for the mode $A_\lambda(t, k)$ until the moment of time when $k = k_S(t^*)$. Starting from this implicitly defined moment of time $t^*(k)$, the conductivities begin to affect the evolution of $A_\lambda(t, k)$. In particular, $A_\lambda(t, k)$ starts to accumulate friction due to σ_E , which suppresses its amplitude by the factor

$$\Delta_k(t) = \exp \left(- \int_{t^*(k)}^t \sigma_E(t') dt' \right), \quad (6.53)$$

In this way, this new definition for $\Theta(t, k)$ in Eq. (6.49) ensures that all the damping effects because of a non-zero conducting medium are still accounted for in a mode's evolution, while also shielding any modes which are in a vacuum state from these effects. Therefore, this represents a simple, but effective, improvement over the previous treatment of conductivities in the literature.

6.2.4 Summary of the effective description

Let us now collect everything we have discussed above to formulate a complete definition of our effective model of Schwinger-induced damping in axion inflation.

The first thing we should do is to consolidate the Ohmic description of the current $\mathbf{J}$ in Eq. (6.19) with the introduction of the window function in Eq. (6.34). Combining the two ideas, we are lead to the following improved definition of the induced current $\mathbf{J}$.

$$\mathbf{J}(t, \mathbf{x}) \equiv \frac{1}{(2\pi)^3} \int d^3\mathbf{y} \left(\sigma_E(t) \mathbf{E}(t, \mathbf{y}) + \sigma_B(t) \mathbf{B}(t, \mathbf{y}) \right) \int d^3\mathbf{k} \Theta(t, k) e^{i\mathbf{k} \cdot (\mathbf{x} - \mathbf{y})}. \quad (6.54)$$

The reader is welcome to check that, with this definition of the induced current, one can indeed retrieve the mode equation in Eq. (6.35) starting from Ampère's law, Eq. (6.16c).

A second point we should address is the definition of the expectation value brackets $\langle \cdot \rangle$ which we use to construct the spatially averaged quantities $\langle \mathbf{E}^2 \rangle$, $\langle \mathbf{B}^2 \rangle$, and $\langle \mathbf{E} \cdot \mathbf{B} \rangle$. Following our discussion of the mode function $A_\lambda(t, k)$ in the previous section, we can define these quantities analogously to the case of pure axion inflation that is, we represent them as the following integrals

$$\langle \mathbf{E}^2 \rangle = \int_0^{k_h(t)} \frac{dk}{k} \frac{k^3}{2\pi^2 a^2} \sum_\lambda |\dot{A}_\lambda(t, k)|^2, \quad (6.55a)$$

$$\langle \mathbf{E} \cdot \mathbf{B} \rangle = \int_0^{k_h(t)} \frac{dk}{k} \frac{k^4}{2\pi^2 a^3} \sum_\lambda \lambda \text{Re}[\dot{A}_\lambda(t, k) A_\lambda^*(t, k)], \quad (6.55b)$$

$$\langle \mathbf{B}^2 \rangle = \int_0^{k_h(t)} \frac{dk}{k} \frac{k^5}{2\pi^2 a^4} \sum_\lambda |A_\lambda(t, k)|^2. \quad (6.55c)$$

The only difference here is the definition of the instability scale $k_h(t)$, which we use as a UV regulator. In the presence of conductivities, we have that the critical scale which separates tachyonically enhanced modes from those that are still in a vacuum-like state is given by k_{crit} in Eq. (6.44). This scale is therefore again a natural choice for a UV regulator. Just as for pure axion inflation, we then generalize the expression in Eq. (6.44) to ensure that it is monotonically increasing even when ξ_{eff} and s_E are not. The appropriate definition that ensures monotonic growth is

$$k_h(t) = \max_{t' \leq t} \left\{ a(t') H(t') \left[|\xi_{\text{eff}}(t')| + \sqrt{\xi_{\text{eff}}^2(t') + s_E^2(t') + s_E(t')} \right] \right\}. \quad (6.56)$$

In the case of no Schwinger damping, we have that $\xi_{\text{eff}} = \xi$, and $s_E = 0$, and this expression for k_h simply reduces to the original one for pure axion inflation in Eq. (3.24).

We can then summarize the entire effective description of axion inflation including charge carriers in terms of the following set of equations;

$$\ddot{\varphi} + 3H\dot{\varphi} + V_{,\varphi} = \frac{\beta}{M_{\text{P}}} \langle \mathbf{E} \cdot \mathbf{B} \rangle, \quad (6.57a)$$

$$\text{div } \mathbf{E} = 0, \quad \text{div } \mathbf{B} = 0, \quad (6.57b)$$

$$\dot{\mathbf{E}} + 2H\mathbf{E} + \sigma_E \bar{\mathbf{E}} - \frac{1}{a} \text{rot } \mathbf{B} + \frac{\beta}{M_{\text{P}}} \dot{\varphi} \mathbf{B} + \sigma_B \bar{\mathbf{B}} = 0, \quad (6.57c)$$

$$\dot{\mathbf{B}} + 2H\mathbf{B} + \frac{1}{a} \text{rot } \mathbf{E} = 0, \quad (6.57d)$$

$$H^2 = \frac{1}{3M_{\text{P}}^2} \left[\frac{1}{2} \dot{\varphi}^2 + V(\varphi) + \frac{1}{2} \langle \mathbf{E}^2 + \mathbf{B}^2 \rangle + \rho_\chi \right], \quad (6.57e)$$

$$\dot{\rho}_\chi + 3H(\rho_\chi + p_\chi) = \sigma_E \langle \mathbf{E} \cdot \bar{\mathbf{E}} \rangle + \sigma_B \langle \mathbf{E} \cdot \bar{\mathbf{B}} \rangle, \quad (6.57f)$$

where we have defined the quantities $\bar{\mathbf{E}}$ and $\bar{\mathbf{B}}$ through

$$\bar{\mathbf{X}}(t, \mathbf{x}) \equiv \frac{1}{(2\pi)^3} \int d^3 \mathbf{y} \mathbf{X}(t, \mathbf{y}) \int d^3 \mathbf{k} \Theta(t, k) e^{i\mathbf{k} \cdot (\mathbf{x} - \mathbf{y})}, \quad \text{for } \mathbf{X} = \mathbf{E}, \mathbf{B}. \quad (6.58)$$

These represent only the fractional part of the electric and magnetic fields, which are subjected to damping effects. The window function which appears here can either be chosen as trivial, $\Theta(t, k) = 1$, as has been the case in the previous literature, or using the improved scale-dependence model defined in Eq. (6.49).

Finally, the conductivities σ_E and σ_B appearing in Eq. (6.57) are to be understood as functions of the expectation values $\langle \mathbf{E}^2 \rangle$, $\langle \mathbf{B}^2 \rangle$, $\langle \mathbf{E} \cdot \mathbf{B} \rangle$ as outlined in Sec. 6.2.2. As we have presented therein, three possible representations for these functions exist: the electric picture ($\sigma_E \neq 0$, $\sigma_B = 0$), the magnetic picture ($\sigma_E = 0$, $\sigma_B \neq 0$), and the mixed picture ($\sigma_E \neq 0$, $\sigma_B \neq 0$). General expressions for these conductivities for all three cases are given in App. C.

The equations in (6.57) encompass all relevant effects of Schwinger pair creation in axion inflation including the production of charge carriers and their subsequent backreaction on the production of gauge fields. Otherwise, the existing structure of the axion inflation model is retained, including the effects of gauge-field backreaction on the inflaton motion.

6.3 Implementation into the GEF

Now that we have seen how the effective model of Schwinger damping in axion inflation can be constructed, the next aspect we want to consider is how the equations of motions in Eq. (6.57) can

be translated into the language of the GEF. Once this is established, we can then use the GEF in the next sections to study the effects of Schwinger damping for a concrete benchmark model.

As the general structure of the equations of motion for the effective model matches that of pure axion inflation, the general strategy that we can employ to construct the GEF equations broadly matches the one we have outlined in chapter 4 for the case of pure axion inflation. Instead of reiterating these individual steps in detail, our focus in this section shall be on the main difference with respect to this construction, namely the question on how we can realize the effects of the scale-dependent damping through $\Theta(t, k)$ within the GEF.

6.3.1 General formulation

The starting point for deriving the GEF equations is once again the definition of the GEF variables $\mathcal{E}^{(n)}$, $\mathcal{B}^{(n)}$ and $\mathcal{G}^{(n)}$ as we have given them in Eq. (4.1). To be explicit, they are given by:

$$\mathcal{E}^{(n)} = \frac{1}{a^n} \langle \mathbf{E} \cdot \text{rot}^n \mathbf{E} \rangle, \quad (6.59a)$$

$$\mathcal{G}^{(n)} = -\frac{1}{2a^n} \langle \mathbf{E} \cdot \text{rot}^n \mathbf{B} + \mathbf{B} \cdot \text{rot}^n \mathbf{E} \rangle, \quad (6.59b)$$

$$\mathcal{B}^{(n)} = \frac{1}{a^n} \langle \mathbf{B} \cdot \text{rot}^n \mathbf{B} \rangle, \quad (6.59c)$$

which we can also represent in terms of gauge-field mode functions, $A_\lambda(t, k)$, as

$$\mathcal{E}^{(n)} = \int_0^{k_h(t)} \frac{dk}{k} \frac{k^{n+3}}{2\pi^2 a^{n+2}} \sum_\lambda \lambda^n |\dot{A}_\lambda(t, k)|^2, \quad (6.60a)$$

$$\mathcal{G}^{(n)} = \int_0^{k_h(t)} \frac{dk}{k} \frac{k^{n+4}}{2\pi^2 a^{n+3}} \sum_\lambda \lambda^{n+1} \text{Re}[\dot{A}_\lambda(t, k) A_\lambda^*(t, k)], \quad (6.60b)$$

$$\mathcal{B}^{(n)} = \int_0^{k_h(t)} \frac{dk}{k} \frac{k^{n+5}}{2\pi^2 a^{n+4}} \sum_\lambda \lambda^n |A_\lambda(t, k)|^2, \quad (6.60c)$$

where we regulate these UV-divergent integrals using $k_h(t)$ defined in Eq. (6.56) (cf. Sec. 3.2).

We can then construct dynamical equations for the GEF variables starting from the system of equations given in Eq. (6.57) and using the procedure outlined in Sec. 4.1. As a result, we obtain the following tower of ODEs describing the evolution of the GEF variables $\mathcal{E}^{(n)}$, $\mathcal{B}^{(n)}$ and $\mathcal{G}^{(n)}$:

$$\dot{\mathcal{E}}^{(n)} + (4+n)H\mathcal{E}^{(n)} + 2\mathcal{G}^{(n+1)} - 2\frac{\beta}{M_P}\dot{\phi}\mathcal{G}^{(n)} + 2\sigma_E\bar{\mathcal{E}}^{(n)} - 2\sigma_B\bar{\mathcal{G}}^{(n)} = S_{\mathcal{E}}^{(n)}, \quad (6.61a)$$

$$\dot{\mathcal{G}}^{(n)} + (4+n)H\mathcal{G}^{(n)} - \mathcal{E}^{(n+1)} + \mathcal{B}^{(n+1)} - \frac{\beta}{M_P}\dot{\phi}\mathcal{B}^{(n)} + \sigma_E\bar{\mathcal{G}}^{(n)} - \sigma_B\bar{\mathcal{B}}^{(n)} = S_{\mathcal{G}}^{(n)}, \quad (6.61b)$$

$$\dot{\mathcal{B}}^{(n)} + (4+n)H\mathcal{B}^{(n)} - 2\mathcal{G}^{(n+1)} = S_{\mathcal{B}}^{(n)}. \quad (6.61c)$$

Here, we implicitly assume that σ_E and σ_B are given as external functions of the mean electric and magnetic background fields in each Hubble patch, i.e., that they are themselves given as functions of $\mathcal{E}^{(n)}$, $\mathcal{B}^{(n)}$ and $\mathcal{G}^{(n)}$ (cf. Sec. 6.2.2 and Ref. [128]).

Just as for the GEF in pure axion inflation, the source terms on the right-hand side of these equations originate from the time-dependent UV regulator $k_h(t)$. They are given by

$$S_{\mathcal{E}}^{(n)} = \frac{d \ln k_h(t)}{dt} \frac{k_h(t)}{2\pi^2} \left(\frac{k_h(t)}{a(t)} \right)^{n+2} \sum_{\lambda=\pm 1} \lambda^n |\dot{A}_\lambda(t, k_h(t))|^2, \quad (6.62a)$$

$$S_{\mathcal{G}}^{(n)} = \frac{d \ln k_h(t)}{dt} \frac{k_h(t)}{2\pi^2} \left(\frac{k_h(t)}{a(t)} \right)^{n+3} \sum_{\lambda=\pm 1} \lambda^{n+1} \text{Re} [A_\lambda^*(t, k_h(t)) \dot{A}_\lambda(t, k_h(t))], \quad (6.62b)$$

$$S_{\mathcal{B}}^{(n)} = \frac{d \ln k_h(t)}{dt} \frac{k_h(t)}{2\pi^2} \left(\frac{k_h(t)}{a(t)} \right)^{n+4} \sum_{\lambda=\pm 1} \lambda^n |A_\lambda(t, k_h(t))|^2, \quad (6.62c)$$

where $A_\lambda(t, k_h(t))$ is the gauge-field mode function whose momentum equals the threshold momentum $k_h(t)$ at this moment of time, i.e., $k = k_h(t)$. To evaluate these mode functions, we can again rely on an analytical solution to the damped mode equation in Eq. (6.35), which is valid if all quantities appearing in this equation evolve adiabatically slowly, i.e., ξ_{eff} , s_E and H . These solutions are again given in terms of Whittaker functions. Their precise form is given in App. A.

The new ingredient with respect to our previous formulation of the GEF is the appearance of the quantities $\bar{\mathcal{E}}^{(n)}$, $\bar{\mathcal{B}}^{(n)}$ and $\bar{\mathcal{G}}^{(n)}$ in Eq. (6.61). These are defined as

$$\bar{\mathcal{E}}^{(n)} = \frac{1}{a^n} \langle \mathbf{E} \cdot \text{rot}^n \bar{\mathbf{E}} \rangle, \quad (6.63a)$$

$$\bar{\mathcal{G}}^{(n)} = -\frac{1}{2a^n} \langle \mathbf{E} \cdot \text{rot}^n \bar{\mathbf{B}} + \mathbf{B} \cdot \text{rot}^n \bar{\mathbf{E}} \rangle, \quad (6.63b)$$

$$\bar{\mathcal{B}}^{(n)} = \frac{1}{a^n} \langle \mathbf{B} \cdot \text{rot}^n \bar{\mathbf{B}} \rangle, \quad (6.63c)$$

where $\bar{\mathbf{E}}$ and $\bar{\mathbf{B}}$ are the electric and magnetic fields convoluted with the window function $\Theta(t, k)$, as we have defined them in Eq. (6.58). Expressed as integrals, these new variables are given by

$$\bar{\mathcal{E}}^{(n)} = \int_0^{k_h(t)} \frac{dk}{k} \frac{k^{n+3}}{2\pi^2 a^{n+2}} \Theta(t, k) \sum_{\lambda} \lambda^n |\dot{A}_\lambda(t, k)|^2, \quad (6.64a)$$

$$\bar{\mathcal{G}}^{(n)} = \int_0^{k_h(t)} \frac{dk}{k} \frac{k^{n+4}}{2\pi^2 a^{n+3}} \Theta(t, k) \sum_{\lambda} \lambda^{n+1} \text{Re} [\dot{A}_\lambda(t, k) A_\lambda^*(t, k)], \quad (6.64b)$$

$$\bar{\mathcal{B}}^{(n)} = \int_0^{k_h(t)} \frac{dk}{k} \frac{k^{n+5}}{2\pi^2 a^{n+4}} \Theta(t, k) \sum_{\lambda} \lambda^n |A_\lambda(t, k)|^2. \quad (6.64c)$$

These terms are evidently necessary to account for the scale-dependent damping effect which we have included into our effective modelling of Schwinger damping in the previous section, which otherwise does not appear in the quantities $\mathcal{X}^{(n)}$ defined in Eq. (6.60). Looking at the definition of the $\mathcal{X}^{(n)}$'s and considering how they appear in the evolution equation in Eq. (6.61), we can see that they are relevant for regulating the damping effects felt by the $\mathcal{X}^{(n)}$'s through σ_E and σ_B . The details of how this occurs in practice evidently depend on the choice of $\Theta(t, k)$. Also, we need to derive an independent evolution equation for the $\bar{\mathcal{X}}^{(n)}$'s to fully describe the dynamical evolution of our new model. This is again subject to the choice of $\Theta(t, k)$.

6.3.2 Capturing scale-dependent effects

In Sec. 6.2.3, we presented two possible choices for $\Theta(t, k)$. The simplest choice was $\Theta(t, k) = 1$, which corresponds to the case where Schwinger damping affects gauge modes indiscriminately of their wavenumber, k . The other choice is to model $\Theta(t, k)$ as in Eq. (6.49), i.e., with damping being only accounted for on modes whose wavelengths are larger than the characteristic separation between charge carriers in the conductive medium at a given moment of time. The first choice of $\Theta(t, k)$ is simple to implement into the GEF, and we shall briefly discuss it first. The second choice proves more challenging, and amounts to defining a second tower of GEF equations for the quantities $\bar{\mathcal{X}}^{(n)}$. In addition to these two possibilities, we also introduce a third choice for $\Theta(t, k)$ which is designed to capture scale-dependent effects, but forgoes the need for a second set of GEF equations.

No scale dependence

We begin with the simplest implementation, which is the case where Schwinger damping affects all gauge modes equally on all scales, i.e., $\Theta(t, k) = 1$. As evident from Eqs. (6.60) and (6.64), this case immediately implies that $\bar{\mathcal{X}}^{(n)} = \mathcal{X}^{(n)}$ for $\mathcal{X} = \mathcal{E}, \mathcal{G}$, or $\mathcal{B}$. This version of the GEF is equivalent to the one which existed in the literature prior to our work, e.g., in Ref. [128]. The only difference is that these authors worked in the electric picture, i.e., $\sigma_B = 0$. Starting from their formulation, an additional magnetic conductivity can be implemented via the replacement $(\beta/M_P)\dot{\phi} \rightarrow ((\beta/M_P)\dot{\phi} + \sigma_B)$ in Eq. (6.61).

Damping on scales $k \leq k_S(t)$

Next, let us consider our second choice for $\Theta(t, k)$, which we proposed in Sec. 6.2.3 as a more physically motivated modelling of Schwinger damping,

$$\Theta(t, k) = \theta(k_S(t) - k)\theta(k_S(t) - a(t)H(t)), \quad (6.65)$$

with k_S defined as in Eq. (6.50). In total, there are then three characteristic scales in our model, which are relevant for the evolution of the GEF system. The instability scale k_h , which separates tachyonically enhanced modes from those which are not, the Schwinger damping scale k_S , which separates damped modes from those that are not, and the Hubble scale, aH , which determines the overall efficiency of Schwinger pair production based on the hierarchy between k_S and aH .

In this version of the GEF, the typical time evolution of the GEF variables $\mathcal{X}^{(n)}$ is then dependent on the hierarchy between $k_h(t)$, $k_S(t)$, and $a(t)H(t)$ at a given time t . Let us therefore consider a typical time evolution of these three characteristic scales, where we focus on what their hierarchy entails for the evolution of a mode function $A_\lambda(t, k)$ depending on its wavenumber, and how this affects $\mathcal{X}^{(n)}$. To guide this discussion, we have also included a visualization of this chronology in Fig. 6.1(a). To start out, we consider a situation of weak gauge-field production, in which Schwinger pair creation is inefficient. This entails that $k_S(t) < a(t)H(t)$. At this stage, no modes $A_\lambda(t, k)$ are subjected to any damping effects. Consequently, all modes that contribute to $\mathcal{X}^{(n)}$ are not damped. A representative mode in this stage of evolution is the one labelled k_1 in Fig. 6.1. As time progresses, electromagnetic fields grow, and Schwinger pair production becomes efficient, i.e., $k_S(t) > a(t)H(t)$. Now, a mode of wavenumber k will experience damping as soon as $k \leq k_S(t)$. However, as long as $k_S(t) < k_h(t)$, any mode will first experience the tachyonic instability before becoming damped. This is illustrated by the mode labeled k_2 in Fig. 6.1(a). Consequently, only some modes contributing to $\mathcal{X}^{(n)}$ are subjected to the damping medium. Only when $k_h(t) \leq k_S(t)$ will a mode experience damping before being subjected to the tachyonic

instability. This is illustrated by the mode labeled k_3 in Fig. 6.1(a). In this case, all modes entering $\mathcal{X}^{(n)}$ are affected by Schwinger damping.

The fact that initially only a fraction of modes contributing to $\mathcal{X}^{(n)}$ are dampened is precisely reflected in the quantities $\bar{\mathcal{X}}^{(n)}$ which we defined in Eq. (6.64). In the case at hand, they are given by

$$\bar{\mathcal{E}}^{(n)}(t) = \int_0^{k_{\text{cut}}(t)} \frac{dk}{k} \frac{k^{n+3}}{2\pi^2 a^{n+2}} \sum_{\lambda} \lambda^n |\dot{A}_{\lambda}(t, k)|^2 \theta_S(t) \equiv \theta_S(t) \tilde{\mathcal{E}}^{(n)}(t), \quad (6.66a)$$

$$\bar{\mathcal{G}}^{(n)}(t) = \int_0^{k_{\text{cut}}(t)} \frac{dk}{k} \frac{k^{n+4}}{2\pi^2 a^{n+3}} \sum_{\lambda} \lambda^{n+1} \text{Re}[\dot{A}_{\lambda}(t, k) A_{\lambda}^*(t, k)] \theta_S(t) \equiv \theta_S(t) \tilde{\mathcal{G}}^{(n)}(t), \quad (6.66b)$$

$$\bar{\mathcal{B}}^{(n)}(t) = \int_0^{k_{\text{cut}}(t)} \frac{dk}{k} \frac{k^{n+5}}{2\pi^2 a^{n+4}} \sum_{\lambda} \lambda^n |A_{\lambda}(t, k)|^2 \theta_S(t) \equiv \theta_S(t) \tilde{\mathcal{B}}^{(n)}(t), \quad (6.66c)$$

with $k_{\text{cut}}(t) = \min(k_S(t), k_h(t))$ and $\theta_S(t) = \theta(k_S(t) - a(t)H(t))$. We can note that the $\bar{\mathcal{X}}^{(n)}$'s account only for modes which have both already undergone Schwinger damping and tachyonic enhancement, in contrast to $\mathcal{X}^{(n)}$ in Eq. (6.60), which accounts for all modes that have been enhanced in their pre-history. Thus, as long as $k_S(t) \leq k_h(t)$, we have that $|\bar{\mathcal{X}}^{(n)}/\mathcal{X}^{(n)}| < 1$. Later, once $k_S(t) \geq k_h(t)$, the two are equal, $\bar{\mathcal{X}}^{(n)} = \mathcal{X}^{(n)}$.

In the last line of Eq. (6.66), we have also defined the terms $\tilde{\mathcal{X}}^{(n)}$. For these auxiliary variables, one can define a new set of GEF equations, which can be solved alongside the ones for $\mathcal{X}^{(n)}$ in Eq. (6.61) to determine the evolution of $\tilde{\mathcal{X}}^{(n)}$ while $k_S(t) < k_h(t)$. These ODEs can be constructed starting from the integral expression for $\tilde{\mathcal{X}}^{(n)}$ in Eq. (6.66), and using the fact that the mode functions $A_{\lambda}(t, k)$ inside the integral evolve according to Eq. (6.35). The resulting equations read

$$\dot{\tilde{\mathcal{E}}}^{(n)} + [(4+n)H + 2\sigma_E] \tilde{\mathcal{E}}^{(n)} + 2\tilde{\mathcal{G}}^{(n+1)} - 2\frac{\beta}{M_P} \dot{\phi} \tilde{\mathcal{G}}^{(n)} - 2\sigma_B \tilde{\mathcal{G}}^{(n)} = \tilde{S}_{\mathcal{E}}^{(n)}, \quad (6.67a)$$

$$\dot{\tilde{\mathcal{G}}}^{(n)} + [(4+n)H + \sigma_E] \tilde{\mathcal{G}}^{(n)} - \tilde{\mathcal{E}}^{(n+1)} + \tilde{\mathcal{B}}^{(n+1)} - \frac{\beta}{M_P} \dot{\phi} \tilde{\mathcal{B}}^{(n)} - \sigma_B \tilde{\mathcal{B}}^{(n)} = \tilde{S}_{\mathcal{G}}^{(n)}, \quad (6.67b)$$

$$\dot{\tilde{\mathcal{B}}}^{(n)} + (4+n)H \tilde{\mathcal{B}}^{(n)} - 2\tilde{\mathcal{G}}^{(n+1)} = \tilde{S}_{\mathcal{B}}^{(n)}. \quad (6.67c)$$

Once again, the time-dependent upper integration limit $k_S(t)$ in Eq. (6.66) leads to boundary terms, namely;

$$\tilde{S}_{\mathcal{E}}^{(n)} = \frac{d \ln k_S(t)}{dt} \frac{k_S(t)}{2\pi^2} \left(\frac{k_S(t)}{a(t)} \right)^{n+2} \sum_{\lambda=\pm 1} \lambda^n |\dot{A}_{\lambda}(t, k_S(t))|^2, \quad (6.68a)$$

$$\tilde{S}_{\mathcal{G}}^{(n)} = \frac{d \ln k_S(t)}{dt} \frac{k_S(t)}{2\pi^2} \left(\frac{k_S(t)}{a(t)} \right)^{n+3} \sum_{\lambda=\pm 1} \lambda^{n+1} \text{Re} [A_{\lambda}^*(t, k_S(t)) \dot{A}_{\lambda}(t, k_S(t))], \quad (6.68b)$$

$$\tilde{S}_{\mathcal{B}}^{(n)} = \frac{d \ln k_S(t)}{dt} \frac{k_S(t)}{2\pi^2} \left(\frac{k_S(t)}{a(t)} \right)^{n+4} \sum_{\lambda=\pm 1} \lambda^n |A_{\lambda}(t, k_S(t))|^2. \quad (6.68c)$$

These boundary terms, $\tilde{S}_{\mathcal{X}}^{(n)}$, account for the fact that, over time, more modes experience the damping conductive medium, and start to contribute to $\tilde{\mathcal{X}}^{(n)}$. This is analogous to the boundary

terms $S_{\mathcal{X}}^{(n)}$ for $\mathcal{X}^{(n)}$, which account for the influx of new modes being tachyonically enhanced as time progresses.

Let us briefly also summarize how the mode functions $A_{\lambda}(t, k)$ entering either into $S_{\mathcal{X}}^{(n)}$ or $\tilde{S}_{\mathcal{X}}^{(n)}$ are subjected to damping depending on the hierarchical structure between the scales k_S , k_h , and aH . We start with those relevant for $S_{\mathcal{X}}^{(n)}$, which, as a reminder, are modes with wavenumber $k = k_h(t)$ at a given time t . Initially, these modes are not damped as long as $k_S(t) \leq k_h(t)$. Only once $k_S(t) \geq k_h(t)$ are the modes crossing the instability scale already damped. This is different for modes that enter into $\tilde{S}_{\mathcal{X}}^{(n)}$. These are modes with wavenumber $k = k_S(t)$ at time t . Per definition, they have never been subjected to damping at any time before t . For practical applications, one can again express the modes entering $S_{\mathcal{X}}^{(n)}$ and $\tilde{S}_{\mathcal{X}}^{(n)}$ in terms of approximate analytical expression involving Whittaker functions. They are summarized in App. A.

Damping on scales $k \leq k_h(t)$

The model we have just described, although based on sound physical motivation, comes with significant challenges in the numerical implementation, since it requires consistently tracking the time evolution of the three scales k_S , k_h , and aH , including the necessity for a second tower of GEF equations for the quantities $\bar{\mathcal{X}}^{(n)}$. In light of this, we also propose a third variant, which allows for an easier numerical implementation, while still capturing the principle idea of the previous model, namely that Schwinger damping should not affect modes of arbitrarily small wavenumbers. In this model, the function $\Theta(t, k)$ has the form

$$\Theta(t, k) = \theta(k_h(t) - k)\theta(k_S(t) - a(t)H(t)). \quad (6.69)$$

This choice of $\Theta(t, k)$ evokes that, as soon as the produced gauge field is sufficiently strong for the Schwinger effect to be in operation, $a(t)H(t) < k_S(t)$, every mode starts feeling Schwinger damping at the same time when it undergoes the tachyonic instability. In other words, modes that are deep inside the horizon ($k \gg aH$) are never damped, while *all* modes outside the instability horizon ($k \leq k_h(t)$) feel the presence of the conducting medium. To illustrate this idea, and how it contrasts to the previous model, see Fig. 6.1(b).

In this simplified case, we always have

$$\bar{\mathcal{X}}^{(n)}(t) = \mathcal{X}^{(n)}(t)\theta(k_S(t) - a(t)H(t)) \quad (6.70)$$

for $\mathcal{X} = \mathcal{E}$, $\mathcal{G}$, and $\mathcal{B}$ because any mode contributing to the $\mathcal{X}^{(n)}$ feels the damping medium as soon as $a(t)H(t) < k_S(t)$. Furthermore, we only have a single set of boundary terms, namely $S_{\mathcal{X}}^{(n)}$. Since we identify the moment of time when a mode crosses the instability scale as the same moment of time when it becomes damped, the mode function $A_{\lambda}(t, k)$ entering these boundary terms can always be treated as having never been subjected to damping effects.

As we shall see in the next section, this simpler model yields comparable results to the previous one, where Schwinger damping is accounted for on scales $k \leq k_S$.

6.3.3 Truncation and error estimation

For each of the three GEF variants which we have discussed above, the infinite tower of differential equations for the variables $\mathcal{X}^{(n)}$ in Eq. (6.61) needs to be analytically closed at some order n_{cut} to arrive at a finite number of ordinary differential equations that can be solved numerically. This truncation scheme is precisely the same as the one which we outlined in chapter 4, that is, we use Eq. (4.13) to approximately relate quantities $\mathcal{X}^{(n_{\text{cut}}+1)}$ to quantities with $n < n_{\text{cut}}$.

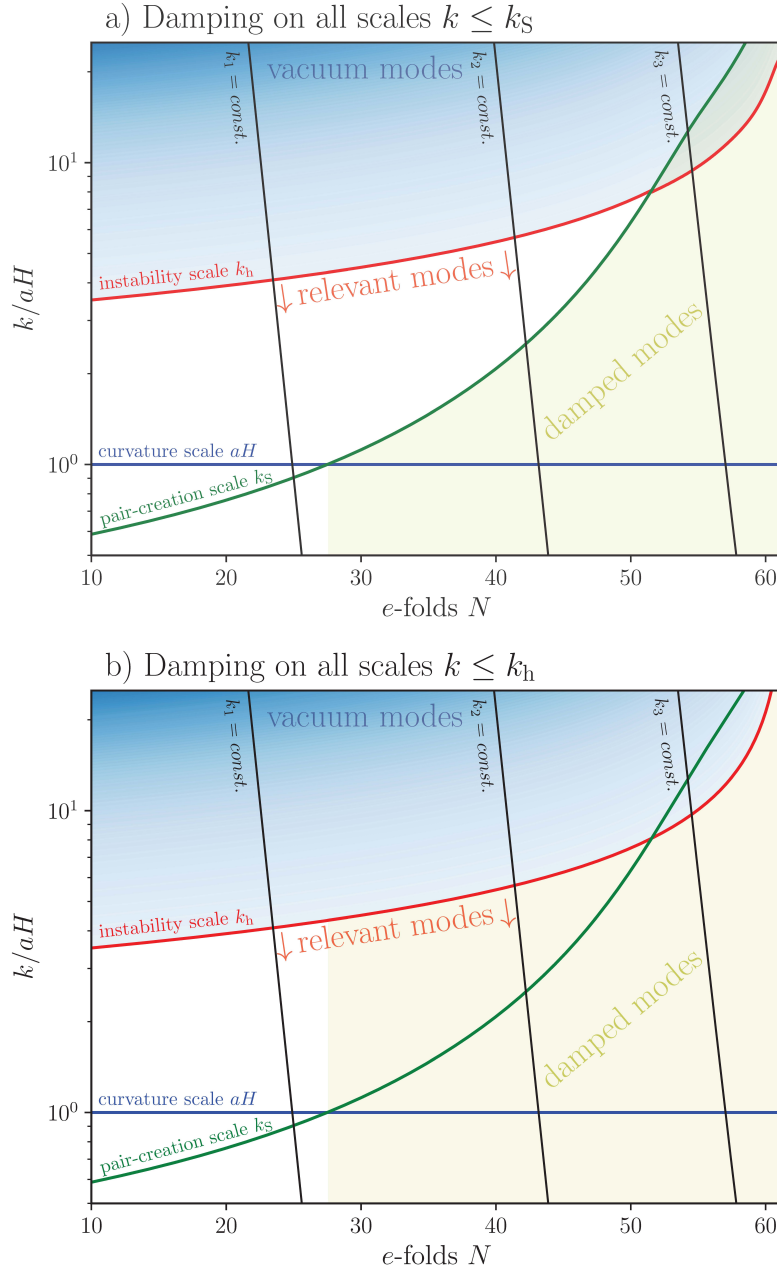


FIGURE 6.1: Time evolution of the three relevant scales during axion inflation including Schwinger pair production: Hubble scale aH (blue), Schwinger pair-creation scale k_S (green), and tachyonic instability scale k_h (red). The blue-shaded region shows gauge-field modes that are still in a vacuum-like state and have not yet been tachyonically amplified. The green-shaded region denotes the modes that undergo Schwinger damping. **Panel (a)** corresponds to the physically motivated case when the damping occurs for all modes with momenta $k \leq k_S(t)$ once $a(t)H(t) < k_S(t)$. The three tilted black lines correspond to gauge-field modes with qualitatively different evolution histories: the leftmost mode with momentum k_1 is not affected by the conductivity until the Schwinger effect sets in (until it crosses the left vertical boundary of the green-shaded region); the mode in the middle (k_2) first experiences the tachyonic instability and then Schwinger damping; the rightmost mode (k_3) undergoes damping prior to the tachyonic enhancement. **Panel (b)** represents the approximation when the damping is turned on for modes with $k \leq k_h(t)$ once $a(t)H(t) < k_S(t)$. Here, the mode k_1 evolves in the same way as the one in panel (a), while the modes with momenta k_2 and k_3 undergo damping as soon as they become tachyonically unstable.

Analogous considerations apply for the terms $\tilde{\mathcal{X}}^{(n)}$, which appear in the second model where Schwinger damping is applied on scales $k \leq k_S$. The GEF equations for these terms, Eq. (6.67), again couple the quantities $\tilde{\mathcal{X}}^{(n)}$ for a fixed n to quantities of order $n + 1$. To truncate these GEF equations, we apply the analogous considerations that lead to Eq. (4.13) for the $\mathcal{X}^{(n)}$'s: the integrals representing these quantities in Eq. (6.66) are dominated for large n by their values at the upper integration boundary, this time k_S instead of k_h . The analytical closing condition for $\tilde{\mathcal{X}}^{(n)}$ is then

$$\tilde{\mathcal{X}}^{(n+1)} \approx \sum_{l=1}^L (-1)^{l-1} \binom{L}{l} \left[\frac{k_S}{a} \right]^{2l} \tilde{\mathcal{X}}^{(n+1-2l)}, \quad \tilde{\mathcal{X}} = \tilde{\mathcal{E}}, \tilde{\mathcal{G}}, \tilde{\mathcal{B}}, \quad (6.71)$$

which is the same as Eq. (4.13), but replacing k_h by k_S .

The numerical integration strategy again follows the one outlined in chapter 4. The ODEs describing the GEF system can still be solved using standard Runge–Kutta algorithms. Initial conditions need to be defined for all dynamical quantities, which now also encompasses charge carriers, in particular their energy density, ρ_χ . For instance, for a system which starts out with weak electromagnetic fields and no initial abundance of charge carriers, $\rho_\chi(0) = 0$.

Once the GEF system has been integrated, one can use the resulting solution as basis to also determine the spectrum of gauge-mode functions $A_\lambda(t, k)$ by solving their mode equation. In the presence of charge carriers, this mode equation is given by Eq. (6.35), with $\Theta(t, k)$ set according to the specific modelling assumptions. Based on this spectrum of mode functions, one can again perform an internal error estimation of the GEF system as in Eq. (4.14), that is, by comparing the quantities $\mathcal{X}^{(0)}$ as obtained by the GEF to the same quantities obtained by integrating the gauge-mode spectrum using the integral representations of $\mathcal{X}^{(0)}$ in Eq. (6.60). Similarly, in case this comparison does not yield satisfactory results, one can apply the self-correction algorithm outlined in chapter 4 to obtain a convergent solution.

At the time of writing, the version of our publicly available Python package, `GEFF v0.1` [7], is able to handle both the basic GEF version, where Schwinger damping is modelled as scale-independent $\Theta(t, k) = 1$, and the more advanced version, where damping effects are accounted for at scales $k \leq k_h(t)$ defined in Eq. (6.69). The full model, where Schwinger damping is accounted for at scales $k \leq k_S(t)$ is currently implemented in a private version of the `GEFF`, which we plan on making publicly available for `GEFF v1.0`.

6.4 Case study: Axion inflation coupled to hypercharge

Having established how our revised effective model can be incorporated into the GEF, we can now consider its practical application to a concrete benchmark scenario. Following the approach of Ref. [128], we here want to consider the case where the gauge field A_μ , which is coupled to the inflaton field, is identified with the $U(1)_Y$ hypercharge field of the Standard Model. This means, that all SM fermions are coupled to A_μ , as well as the SM Higgs field, and all of them can, in principle, be created via the Schwinger effect during axion inflation.

In what follows, we first discuss how this concrete assumption regarding the gauge sector can be reflected in the effective model that we devised in Sec. 6.2. In this, we follow the ideas outlined in Ref. [128]. After this, we perform an in-depth analysis regarding the dynamical effects of Schwinger damping in this model. First, we compare how the dynamics of axion inflation differ when accounting for the Schwinger effect versus when the Schwinger effect is unimportant, i.e., we compare fermionic axion inflation to pure axion inflation. Then, we perform a comparison between our new method for effectively modelling the Schwinger effect, i.e., including scale-dependent damping and using the mixed-picture representation, to prior effective modelling assumptions that

exist in the literature, i.e., no scale-dependent damping and using either the electric or magnetic picture to represent the induced Schwinger current $\mathbf{J}$. Finally, we discuss a novel dynamical regime of gauge-field backreaction in axion inflation, which arises when sufficiently strong gauge-field production is accompanied by efficient Schwinger pair creation.

6.4.1 Modelling the production of charged particles

In Sec. 6.2, we have given a general framework to account for the effect of charged fields that can be produced during axion inflation via the Schwinger effect. For this, we have implicitly assumed a single species of charge carriers, χ . In the model we are presently interested in, we have that multiple fields χ_i are coupled to the gauge field A_μ , namely, every SM fermion and the Higgs field Φ . Let us therefore briefly consider how the framework we have developed so far can account for multiple charge carriers.

Information on charge carriers enters in two places in our effective model: the conductivities σ_E and σ_B , and in the continuity equation for the energy density ρ_χ , Eq. (6.18). For simplicity, in what follows, we assume that interactions in the charged sector are negligible. Furthermore, to give concrete expressions, we always assume that we are working in the mixed picture, defined in Eq. (6.19). Expressions for the electric and magnetic picture can be analogously determined by relying on the formulae given in App. C.

Let us first consider the conductivities. For multiple charge carriers, we have a total Schwinger-induced current $\mathbf{J}$, which is composed of the individual currents for each charged field χ_i , i.e., $\mathbf{J} = \sum_i \mathbf{J}_i$. In the collinear frame, each current $\mathbf{J}_i$ is aligned with the electric and the magnetic field, such that the magnitude of the total current is given by the sum over the magnitudes of the individual currents

$$|\mathbf{J}'| = \sum_i |\mathbf{J}'_i| = \sum_i \frac{(e|Q_i|)^3}{6\pi^2 H} E' |B'| \psi_i \left(\frac{\pi |B'|}{E'} \right) \exp \left(-\frac{\pi m_i^2}{e|Q_i|E'} \right), \quad (6.72)$$

where the prime ($'$) denotes quantities in the collinear frame. The total current therefore depends on the masses m_i , and the charges Q_i of each field χ_i , while their spin representation defines the function $\psi_i(x)$, i.e., $\psi_i(x) = \coth x$ for a fermion, or $\psi_i(x) = 1/(2 \sinh x)$ for a charged scalar. From Eq. (6.72), one may then compute the total induced conductivities $\sigma_{E/B}$ using, e.g., the mixed-picture representation in Eq. (6.29).

Similarly, we can generalize the continuity equation for the energy density of charge carriers ρ_χ in Eq. (6.18) to account for multiple fields. The obvious generalization is that the total energy density is the sum over the individual energy densities $\rho_\chi = \sum_i \rho_\chi^i$, and similarly for the total pressure, $p_\chi = \sum_i p_\chi^i$. The pressure p_χ^i for each species χ_i needs to be specified via an effective equation of state parameter; $p_\chi^i = w_\chi^i \rho_\chi^i$. If all particles are massless, we have $w_\chi^i = 1/3$, thus $p_\chi = \rho_\chi/3$. Once again, we are neglecting any interactions between charge carrier species for this argument to hold.

After these general remarks, let us consider the physics model we are interested in more closely. If A_μ is the hypercharge field, all SM fermions and the Higgs field Φ can prospectively be produced via Schwinger pair creation in our model. The first thing we should then consider is what the masses of these fields are. Following the ideas outlined in Ref. [128], we focus on the case where the electroweak symmetry is unbroken during inflation, such that all SM fermions are massless. Looking at Eq. (6.72), this implies that the individual Schwinger current $\mathbf{J}_i$ for each fermion species is not subjected to an exponential suppression by a mass term, and their production through the hypercharge field A_μ can be highly efficient. For this assumption to be valid, the Higgs needs to be stabilized at the origin of field space. One possibility to ensure that

this is true during inflation is to include a non-minimal coupling between the Higgs field and Ricci curvature, $\mathcal{L} \supset \xi_R |\Phi|^2 R$. This, however, also entails that the Higgs field is heavy during inflation, $m_\Phi^2 \propto \xi_R H^2$. To be consistent, we therefore do not account for the possibility of Higgs production via the Schwinger effect, since the associated current would be exponentially suppressed by this large mass.

Building on these assumptions, we therefore have that the total Schwinger-induced current in our model is composed of the sum over the currents for all SM fermions, each weighted by the cube of their hypercharge (see Eq. (6.72)). Concretely, this implies that the total electric and magnetic conductivities in the mixed picture for this model are given by (see App. C);

$$\sigma_E = \frac{41}{12} \frac{g'(\mu)^3}{6\pi^2 H} \frac{|\langle \mathbf{E} \cdot \mathbf{B} \rangle|}{\sqrt{\Sigma}} \sqrt{\frac{\langle \mathbf{E}^2 - \mathbf{B}^2 \rangle + \Sigma}{\langle \mathbf{E}^2 + \mathbf{B}^2 \rangle + \Sigma}} \coth \left(\pi \sqrt{\frac{\langle \mathbf{B}^2 - \mathbf{E}^2 \rangle + \Sigma}{\langle \mathbf{E}^2 - \mathbf{B}^2 \rangle + \Sigma}} \right), \quad (6.73a)$$

$$\sigma_B = \frac{41}{12} \frac{g'(\mu)^3}{6\pi^2 H} \frac{\langle \mathbf{E} \cdot \mathbf{B} \rangle}{\sqrt{\Sigma}} \sqrt{\frac{\langle \mathbf{B}^2 - \mathbf{E}^2 \rangle + \Sigma}{\langle \mathbf{E}^2 + \mathbf{B}^2 \rangle + \Sigma}} \coth \left(\pi \sqrt{\frac{\langle \mathbf{B}^2 - \mathbf{E}^2 \rangle + \Sigma}{\langle \mathbf{E}^2 - \mathbf{B}^2 \rangle + \Sigma}} \right), \quad (6.73b)$$

with $\Sigma = \sqrt{\langle \mathbf{E}^2 - \mathbf{B}^2 \rangle^2 + 4 \langle \mathbf{E} \cdot \mathbf{B} \rangle^2}$. Note that the pre-factor 41/12 corresponds to the sum of the cubes of the hypercharges for all fermions in the SM, and we have that $\psi(x) = \coth x$, because we are exclusively considering fermions. We also account for the running of the gauge coupling g' ,

$$g'(\mu) = \left([g'(m_Z)]^{-2} + \frac{41}{48\pi^2} \ln \frac{m_Z}{\mu} \right)^{-1/2}, \quad (6.74)$$

which takes into account the contributions of all SM particles to the one-loop beta function (see, for instance, Ref. [423]). Here, $m_Z \approx 91.2$ GeV is the Z -boson mass, $g'(m_Z) \approx 0.35$. We evaluate this expression for the characteristic energy scale for the process of pair creation which we take to be $\mu = \rho_{\text{EM}}^{1/4} = (\langle \mathbf{E}^2 + \mathbf{B}^2 \rangle / 2)^{1/4}$.

Similar considerations also lead us to define the pair-creation scale k_S in our model as (see Eq. (6.50))

$$k_S = a \left(\frac{41}{12} \right)^{1/3} \frac{g'(\mu)}{2^{1/4}} \left[\langle \mathbf{E}^2 + \mathbf{B}^2 \rangle + \sqrt{\langle \mathbf{E}^2 - \mathbf{B}^2 \rangle^2 + 4 \langle \mathbf{E} \cdot \mathbf{B} \rangle^2} \right]^{1/4}. \quad (6.75)$$

Given that we only consider the pair creation of massless fermions, we also have that the total pressure of charge carriers, p_χ entering the evolution equation for the corresponding total energy density ρ_χ in Eq. (6.18) is given by $p_\chi = \rho_\chi / 3$.

This then fully specifies the entire charged sector of our model. The inflationary sector is specified as usual, by defining the inflaton potential shape, and setting the strength of the axion–vector coupling β .

6.4.2 Implications of Schwinger damping

Now that we have established how we can suitably represent the case where the gauge field A_μ is identified with hypercharge, let us consider the effects of Schwinger pair production on the dynamics of axion inflation. To do this, we rely on the GEF implementation outlined in Sec. 6.3.

For this analysis, we need to specify several aspects of the inflationary dynamics. We initialize the inflaton field on a slow-roll trajectory of a quadratic potential, $V(\phi) = m^2 \phi^2 / 2$, starting approximately 60 e -folds before the expected end of slow-roll inflation if backreaction effects are absent, thus $\varphi(0) = 15.55 M_P$ and $\dot{\varphi}(0) = -\sqrt{2/3} m M_P$. Furthermore, we take all gauge-field

expectation values as initially zero, $\mathcal{X}^{(n)}(0) = 0$. Consequently, we also assume that the production of charged particles via the Schwinger effect is initially ineffective, so we initialize $\rho_\chi(0) = 0$. These initial conditions may be simply understood as the generalization of the slow-roll initial conditions we have described in Sec. 4.2.2 for the case where the Schwinger effect is also taken into consideration.

Comparison to pure axion inflation

As a first point of comparison, we consider an axion–vector coupling $\beta = 25$, and the inflaton mass is $m = 6.16 \times 10^{-6} M_{\text{P}}$, such that we may directly compare the results for the present benchmark model to the ones for pure axion inflation given in Sec. 4.2.2 (cf. Fig. 4.1). This is also a common benchmark studied in other works in the literature, e.g., Refs. [109, 128]. As a reminder, in the case of pure axion inflation, this benchmark features an onset of strong gauge-field backreaction and a prolonged phase of inflationary expansion. For this comparison, we rely on the physically most motivated modelling of the Schwinger effect; that is, we consider that damping only affects wavenumbers with $k \leq k_{\text{S}}$ (i.e., $\Theta(t, k)$ is given by Eq. (6.49)), and we work in the mixed-picture representation of the induced Schwinger current.

In Fig. 6.2, we now illustrate how the dynamical evolution of axion inflation changes when including the production of SM fermions. We show two panels; panel (a) depicts the production of electromagnetic fields and fermions over time, while panel (b) gives the relative contributions to the total energy density in the system coming from the inflaton field, the electromagnetic field, and the fermions. Focussing on panel (a), we can see that gauge-field production in this model is initially efficient, and, in particular, the electric energy density, $\rho_E = \langle \mathbf{E}^2 \rangle / 2$ is dominant over the fermionic energy density, ρ_χ . However, this strong electromagnetic field also implies an efficient conversion into fermions via the Schwinger effect, which itself dampens the further generation of gauge fields. Toward the end of inflation, this even leads to ρ_χ becoming dominant with respect to ρ_E , the magnetic energy density, $\rho_B = \langle \mathbf{B}^2 \rangle / 2$, and the Chern–Pontryagin density $\rho_{EB} = \langle \mathbf{E} \cdot \mathbf{B} \rangle / 2$. The consequence of this efficient conversion process can be seen in panel (b) of figure 6.2. Unlike the case of pure axion inflation, we find that the amplification of electromagnetic fields is never sufficient to induce strong-backreaction effects. The dynamics of spacetime are entirely dictated by the inflaton field, whose potential energy drives inflation, and inflation ends once the slow-roll condition breaks down, $\dot{\varphi}^2 \simeq V(\varphi)$. The lower part of panel (a) confirms that these results are under numerical control; the internal comparison between GEF results and mode-by-mode spectra yields an error estimate which remains at the percent level for the entire simulation, with only a mild increase toward the end of inflation. This level of accuracy is entirely sufficient to demonstrate the negligible impact of gauge-field backreaction in this model.

For a realistic model of axion inflation, it can therefore be of paramount importance to account for the Schwinger effect. In particular, a reduced amplitude in electromagnetic fields could limit the potential for magnetogenesis [134], and their imprint of the spectrum of scalar and tensor perturbations may be less pronounced than in the case of pure axion inflation [120]. Meanwhile, the efficiency of fermion production is itself enticing. Their imprint on the spectrum of inflationary perturbations could potentially be equally important as the gauge-field-induced contributions. Additionally, the large amplitude in fermions at the end of inflation can also be relevant for reheating. Indeed, the Schwinger effect has been proposed as a complementary reheating mechanism [124] alongside the direct inflaton decay and the parametric resonance [44, 169, 170].

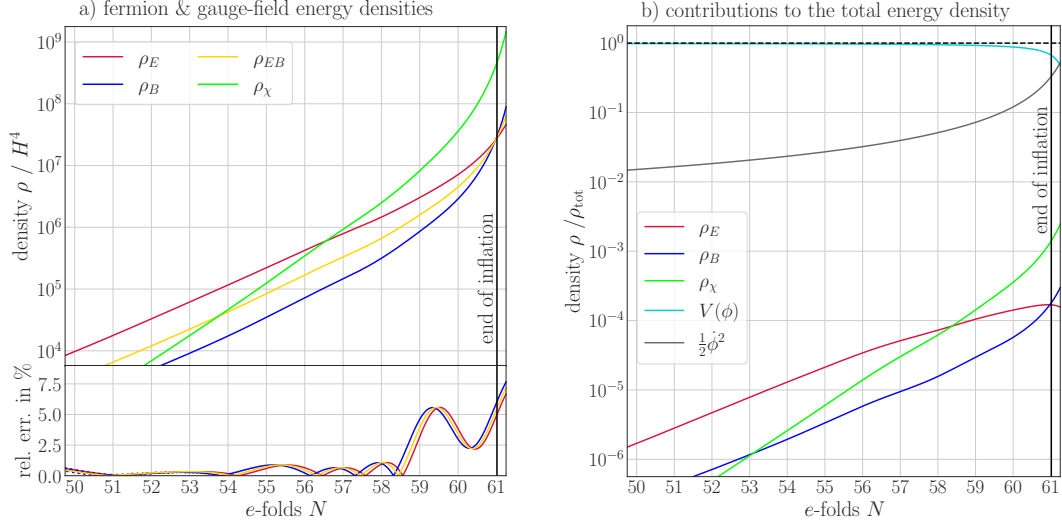


FIGURE 6.2: **Panel (a):** The upper part shows the evolution in e -folds of the electric energy density ρ_E (red), magnetic energy density ρ_B (blue), the Chern–Pontryagin density ρ_{EB} (yellow), and the fermion energy density ρ_χ (green), divided by H^4 . The lower panel shows the relative deviation between the GEF result and the corresponding mode-by-mode solution in momentum space. **Panel (b):** Relative contributions of the electric (red), magnetic (blue) and fermion (green) energy density to the total energy density (dashed black). Also shown are the contributions of the inflaton field to the total energy density stored in kinetic (gray) and potential (cyan) energy. The results in both panels are obtained in the physically motivated model of the Schwinger effect involving the mixed picture for the Schwinger conductivity and accounting for gauge-mode damping on scales $k \leq k_S(t)$. The dimensionless axion–vector coupling is $\beta = 25$ and the inflaton mass is $m = 6.16 \times 10^{-6} M_{\text{P}}$. The results we show here should be compared to the same results for pure axion inflation in Fig. 4.1.

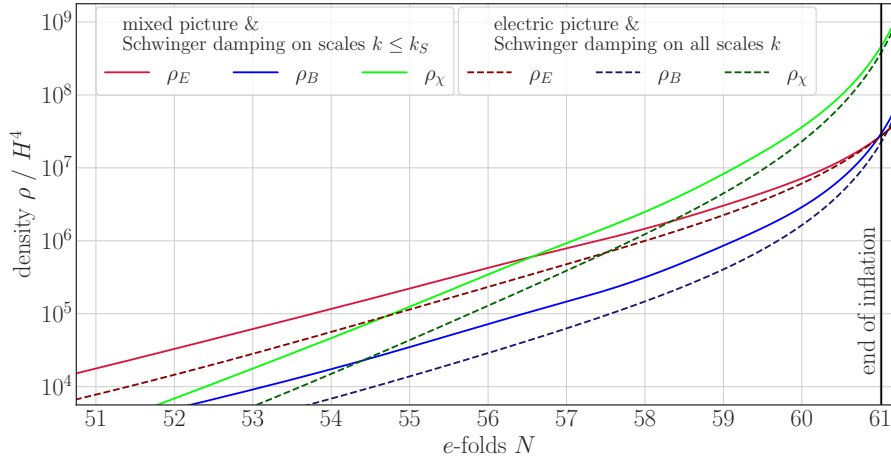


FIGURE 6.3: Evolution in e -folds of the electric (red), magnetic (blue), and fermion (green) energy densities, divided by H^4 . The results are shown for two different modelling assumptions. Dashed lines depict results based on the electric picture and when treating Schwinger damping as scale-independent. This corresponds to the configuration studied in Ref. [128]. Solid lines instead depict results obtained in the physically more motivated model, where Schwinger damping is only accounted for on scales $k \leq k_S(t)$ and conductivities are computed in the mixed picture.

Comparing different modelling assumptions

The results we have shown in Fig. 6.2 are based on our improved modelling of the Schwinger effect, taking into account scale-dependent damping effects and modelling the fermionic conductivities in the mixed picture. We now consider how these results differ with respect to previous approaches in the literature.

First, we compare them to the model of Ref. [128], which assumed the electric picture and did not account for the scale dependence of Schwinger damping (i.e., $\Theta(t, k) = 1$). In Fig. 6.3, we show the time dependence of the gauge-field and fermion energy densities for our improved model, and the model used in Ref. [128]. Qualitatively, the results are in good agreement: the gauge field becomes sufficiently strong to produce fermions, which then dampen gauge-field generation, and ρ_χ becomes dominant toward the end of inflation. However, on the quantitative level, the two results differ significantly up until the very last e -folds of inflation, where all curves (accidentally) converge. Note that we only show the last 10 e -folds of inflation. At earlier times, when the electromagnetic fields are too weak to source fermions via Schwinger pair creation, both models naturally yield equivalent results.

This difference in the predicted amplitude of electromagnetic fields and the abundance of fermions during inflation can be important for phenomenological applications. For instance, this can be relevant when determining how these fields affect the spectrum of primordial perturbations generated during inflation, since the inflationary evolution at a given moment of time t is typically reflected in the power spectra at scales that leave the horizon at this time, $k \sim a(t)H(t)$.

Next, we broaden the scope of our comparison by considering all possible models of the Schwinger effect that we can construct based on our effective model in Sec. 6.2 and the possible GEF implementations in Sec. 6.3. We have presented three possible representations for the induced current (i.e, the electric, the magnetic and the mixed picture), and we have proposed three possible models to describe Schwinger damping which we can implement in the GEF. In total, this gives nine possibilities to model the Schwinger effect in axion inflation. In Figs. 6.4 and 6.5, we show the evolution of the electric (ρ_E , red), magnetic (ρ_B , blue), and fermionic (ρ_χ , green) energy density in addition to the Chern–Pontryagin density (ρ_{EB} , yellow) for each of these nine realizations. These results are shown with a different line style depending on the Ohmic representation of the induced current and a different color shading depending on the scale dependence of Schwinger damping. The two figures, Figs. 6.4 and 6.5, display the same result but arranged differently: in Fig. 6.4, results are ordered such that the time evolution of the densities ρ_E , ρ_B , ρ_{EB} and ρ_χ in the same column are computed using the same model of scale-dependent damping, with each panel showing results for the different Ohmic representations; meanwhile, Fig. 6.5 orders the same results by a shared Ohmic representation in each column while the different models of scale-dependent damping are compared within each panel.

Based on Fig. 6.4, we can easily compare results for the electric, magnetic, and mixed pictures. From this, we can note that results for the electric and mixed pictures are generally in overall good agreement, while the magnetic picture predicts a noticeably smaller amplitude in all four energy densities. To explain the similarity between the electric and the mixed picture it is worth noting that, during axion inflation, electric and magnetic fields are in fact near collinear, implying $\langle \mathbf{E} \cdot \mathbf{B} \rangle^2 \simeq \langle \mathbf{E}^2 \rangle \langle \mathbf{B}^2 \rangle$. Using this to approximate the expressions for σ_E and σ_B in the mixed

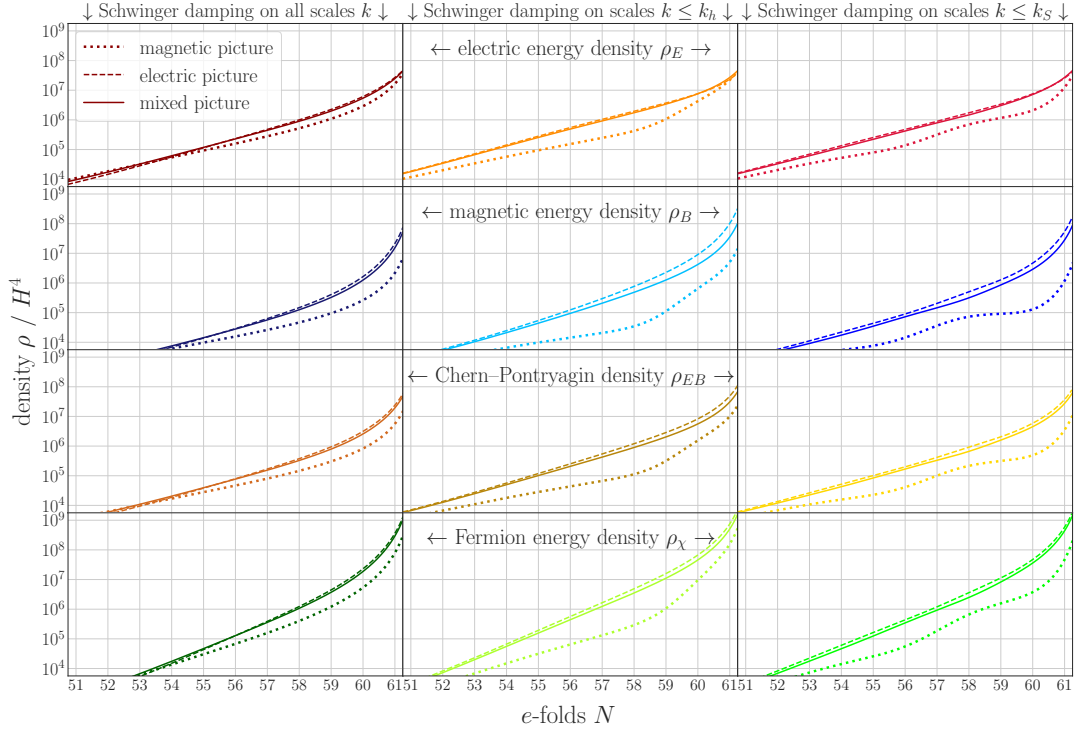


FIGURE 6.4: A comparison between all possible modelling assumptions for Schwinger damping in axion inflation presented in this chapter. We show the evolution in e -folds for the electric energy density ρ_E (first row, shades of red), magnetic energy density ρ_B (second row, shades of blue), Chern–Pontryagin density ρ_{EB} (third row, shades of yellow), and fermion energy density ρ_χ (fourth row, shades of green). All results are computed for $\beta = 25$ and $m = 6.16 \times 10^{-6}$. Different modelling assumptions are distinguished by line style and color shading. The line style distinguishes between the electric picture ($\sigma_E \neq 0, \sigma_B = 0$, dashed lines), the magnetic picture ($\sigma_E = 0, \sigma_B \neq 0$, dotted lines), and the mixed picture ($\sigma_E \neq 0, \sigma_B \neq 0$, solid lines). Schwinger damping can affect scales $k \leq k_S(t)$ (pure red, blue, yellow, or green), all scales (dark shades), or scales $k \leq k_h(t)$ (light shades). In this figure, columns show results based on the same modelling of scale-dependent effects.

picture, Eq. (6.73), one gets

$$\sigma_E \simeq \frac{41}{12} \frac{g'(\mu)^3}{6\pi^2 H} \sqrt{\langle \mathbf{B}^2 \rangle} \frac{\langle \mathbf{E}^2 \rangle}{\langle \mathbf{E}^2 + \mathbf{B}^2 \rangle} \coth \left(\pi \sqrt{\frac{\langle \mathbf{B}^2 \rangle}{\langle \mathbf{E}^2 \rangle}} \right), \quad (6.76a)$$

$$\sigma_B \simeq \text{sgn}(\langle \mathbf{E} \cdot \mathbf{B} \rangle) \frac{41}{12} \frac{g'(\mu)^3}{6\pi^2 H} \sqrt{\langle \mathbf{E}^2 \rangle} \frac{\langle \mathbf{B}^2 \rangle}{\langle \mathbf{E}^2 + \mathbf{B}^2 \rangle} \coth \left(\pi \sqrt{\frac{\langle \mathbf{B}^2 \rangle}{\langle \mathbf{E}^2 \rangle}} \right). \quad (6.76b)$$

We should then account for the observation that the electric energy density is significantly larger than the magnetic one before the end of inflation, thus $\langle \mathbf{E}^2 \rangle / \langle \mathbf{E}^2 + \mathbf{B}^2 \rangle \simeq 1$, while $\langle \mathbf{B}^2 \rangle / \langle \mathbf{E}^2 + \mathbf{B}^2 \rangle \ll 1$. In these limits, Eq. (6.76) implies that the mixed picture is primarily aligned with the electric picture, explaining why they yield comparable results. We can also speculate that the more drastic reduction in the energy densities in the magnetic picture may be a result of the fact that σ_B directly suppresses gauge-field production via the tachyonic enhancement. It is conceivable that this direct impact on gauge-field generation is more efficient than the friction due to σ_E , explaining why the purely magnetic picture features a lower overall amplitude in gauge fields than the other two.

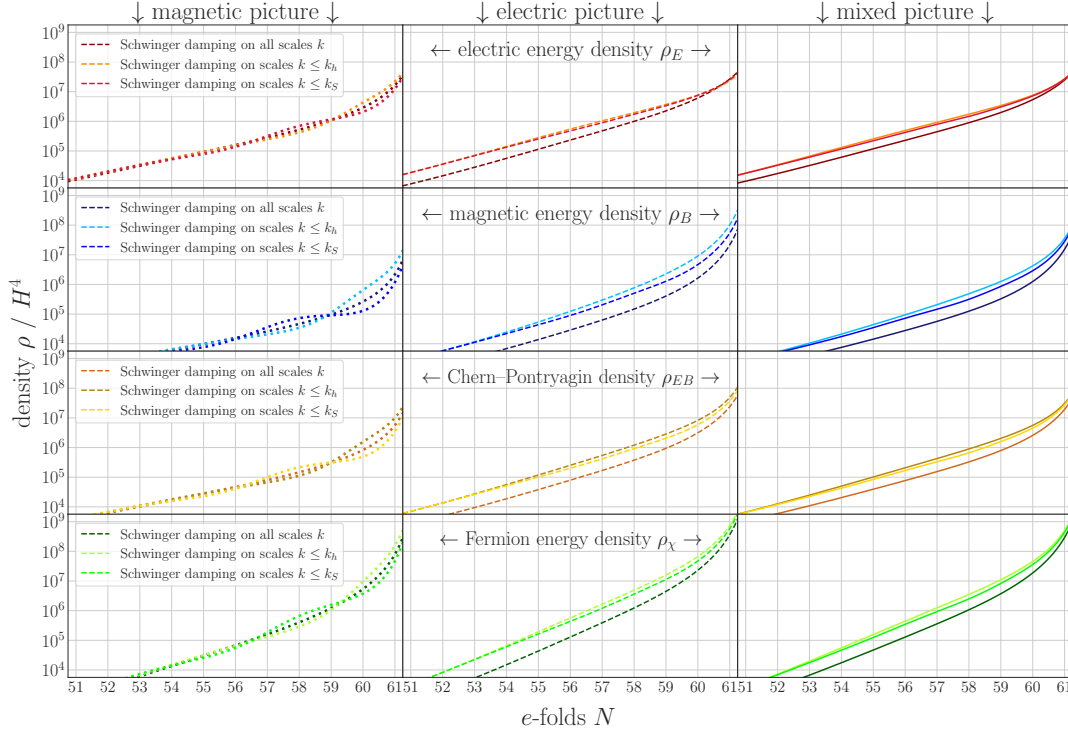


FIGURE 6.5: The same results as in Fig. 6.4, but now each column shows results based on the same Ohmic representation.

To compare different models for the scale dependence of Schwinger damping, we can consider Fig. 6.5. The simplest case where all gauge-field modes are affected by the conductivities irrespective of their momenta, naturally, leads to the strongest suppression of gauge-field generation, and, therefore, tends to underestimate the more refined result where we account for scale-dependent Schwinger damping. The physically motivated model of Schwinger damping on scales $k \leq k_S(t)$ (except for the magnetic picture) predicts larger energy densities for the produced fields by almost one order of magnitude as compared to the case with no scale dependence. A second interesting observation concerns the comparison between the model where we account for damping on scales $k \leq k_S(t)$ and the simplified version of it, where damping only affects gauge modes on scales $k \leq k_h(t)$. We find that both modelling assumptions yield near identical results. In Ref. [5], we noted that the similarity between these two versions of scale-dependent damping can be understood on a quantitative level. We found that, once $k_S > k_h$, i.e., at late times, when gauge-field production and damping effects are both important, the two scales differ at most by an order-one factor. Considering that both scales change exponentially with time, i.e., they change with the scale factor $a(t)$, this small difference therefore does not result in a large discrepancy between the two modelling assumptions. Therefore, given its simpler implementation, the model where Schwinger damping is only effective on scales $k \leq k_h(t)$ is the one we rely on from this point forward.

6.4.3 Fermion-tempered backreaction

We have seen that fermion production during axion inflation comes at the price of gauge-field production. In particular, for the benchmark point $\beta = 25$ and $m = 6.16 \times 10^{-6} M_P$, we have noted that the reduced gauge-field amplitude prevents the onset of gauge-field backreaction, which otherwise would take place if we neglect damping effects from fermions. As a consequence,

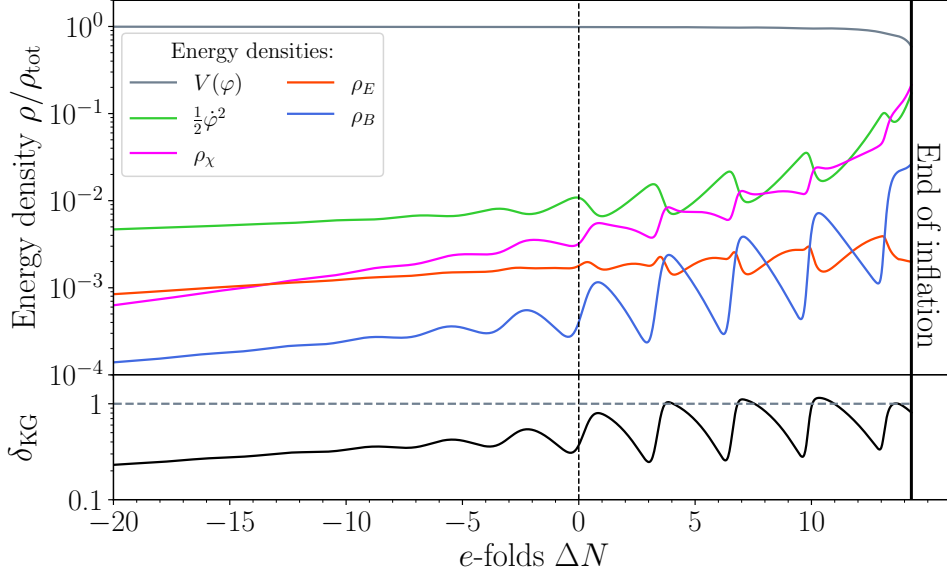


FIGURE 6.6: Results for $\beta = 59$ and $m = 2.5 \times 10^{-5} M_{\text{P}}$. The upper panel shows the evolution of the energy densities $V(\varphi)$ (gray), $\frac{1}{2}\dot{\varphi}^2$ (green), ρ_E (red), ρ_B (blue), and ρ_χ (magenta) normalized to the total energy density $\rho_{\text{tot}} = 3H^2 M_{\text{P}}^2$. The lower panel shows the corresponding evolution of the backreaction parameter δ_{KG} defined in Eq. (3.27). The horizontal axis indicates e -folds with respect to the expected end of slow-roll inflation at $\Delta N = 0$. The evolution of this system is representative for fermion-tempered backreaction.

the results we have considered thus far show none of the hallmark signatures of backreaction, such as oscillations in the inflaton’s kinetic energy, or a prolonged duration of inflation. However, backreaction effects are not generically absent when accounting for Schwinger damping. In Ref. [5], we identified a new region of parameter space where backreaction effects can be present despite the creation of SM fermions inhibiting gauge-field production. For gauge-field backreaction to become relevant, one however needs to consider significantly larger values for the axion–vector coupling, β , than what is necessary in the case of pure axion inflation.

In Fig. 6.6, we show results for a representative parameter-space point for which gauge-field backreaction effects are important even when accounting for Schwinger damping; $\beta = 59$ and $m = 2.5 \times 10^{-5} M_{\text{P}}$. The figure is separated into two panels: the upper one depicts the evolution of the energy densities in the system, while the lower panel shows the evolution of the backreaction parameter δ_{KG} , which quantifies the importance of gauge-field friction in the Klein–Gordon equation (see Eq. (3.27)). Qualitatively, we can recognize similar features in the system’s evolution as in the standard strong-backreaction regime for pure axion inflation; an oscillatory evolution of the inflaton’s kinetic energy density, $\dot{\varphi}^2/2$, which is mimicked by the electromagnetic energy density. The stuttering motion of the inflaton field results in a prolonged duration of inflation.

However, these dynamical effects are not as violent as in the case of pure axion inflation. We should remember that there, strong backreaction can even result in brief inversions of the inflaton’s motion through field space, such that the inflaton field rolls back up its potential hill. Furthermore, strong backreaction in pure axion inflation typically features an overall growth in the electromagnetic energy density, while the inflaton’s kinetic energy is considerably drained, and inflation only ends when $\rho_E \simeq V(\varphi)$. The system whose dynamics we show in Fig. 6.6 displays neither of these effects. Although the inflaton field oscillates, it never deviates far from around the standard slow-roll trajectory, and it is ultimately the inflaton’s kinetic energy which terminates inflation, $\dot{\varphi}^2 \simeq V(\varphi)$. The obvious reason for these differences is that the model in question

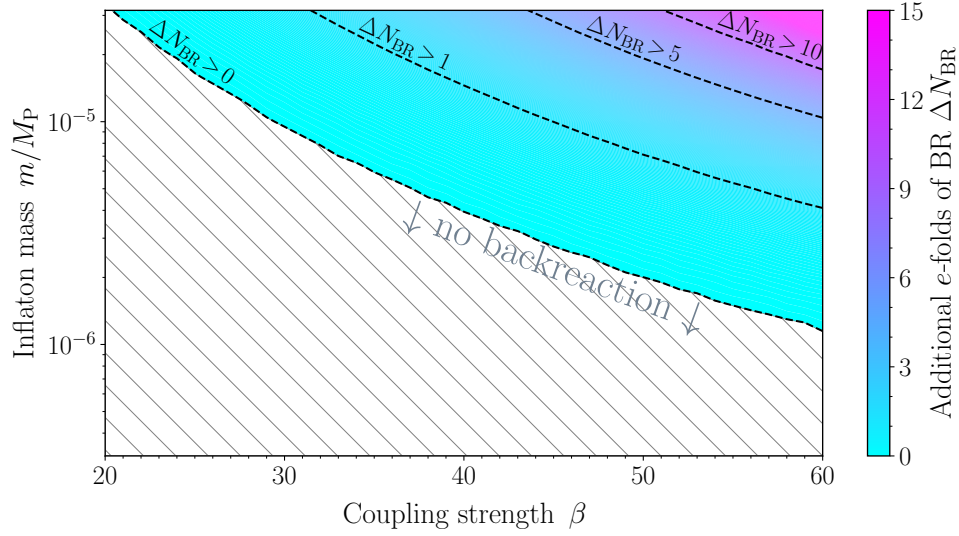


FIGURE 6.7: Results for a parameter scan over the dimensionless axion–vector coupling β and the inflaton mass m when accounting for the production of SM fermions via the Schwinger effect. The colored area shows regions of parameter space where backreaction effects can become important as indicated by a prolonged duration of inflation (in e -folds, ΔN_{BR}) past the expectation based on pure slow-roll dynamics. Fermion-tempered backreaction sets in for large β and m and can extend the duration of inflation by more than 10 e -foldings.

also accounts for fermions. Whenever gauge-field friction grows sufficiently to trigger strong backreaction, fermion production is also increased, and the gauge field is subjected to increased damping effects. This can be seen in Fig. 6.6 by noting that ρ_χ oscillates in phase with ρ_E and ρ_B . Because of this additional damping, δ_{KG} never remains consistently close to unity, and friction effects do not reach similar levels as in pure axion inflation. Consequently, the entire system evolves in a more temperate manner compared to the usual strong-backreaction regime. To emphasize this difference to the standard strong-backreaction regime in pure axion inflation, we refer to this novel dynamical regime as *fermion-tempered backreaction*.

An important difference between fermion-tempered backreaction and the standard strong-backreaction regime is that the former can support a substantially longer period of inflation. For pure axion inflation, we have seen that inflation can be extended by ~ 5 e folds until gauge-field production is efficient enough to terminate inflation (cf., e.g., Figs. 3.3 and 4.1). In the presence of fermions, the oscillatory motion of the system can be sustained for much longer, as we indicate in Fig. 6.7. Therein, we show the number of additional e -folds of inflation due to backreaction as a function of our model parameters, β and m . The displayed results are obtained by performing a parameter scan over β and m using the GEF. We find that, for large couplings and large masses, inflation can be extended by up to 20 e -folds. Viewing an extended duration of inflation as a smoking-gun signature of backreaction, Fig. 6.7 also allows us to identify that fermion-tempered backreaction in our model occurs for large couplings, $\beta \gtrsim 40$, and large values of the inflaton mass, $m \gtrsim 3 \times 10^{-6} M_{\text{P}}$.

It should be noted that the results shown in this section have been derived for the simplified model of Schwinger damping, wherein we account for damping effects for all gauge-field modes with wavenumbers $k \leq k_{\text{h}}$. However, we have checked explicitly that fermion-tempered backreaction also occurs in the more refined model where Schwinger damping can only affect gauge modes on scales $k \leq k_{\text{S}}$. For more details on this comparison, we refer the reader to App. A of our original article, Ref. [5].

6.5 Summary and conclusion

In this chapter, we studied the dynamical interplay between Schwinger pair creation and gauge-field production during axion inflation. As we have explicitly seen, the continuous creation of charge carriers from the vacuum can be viewed as inducing a conductive medium which dampens the further generation of gauge fields. These effects can be captured by means of an effective treatment of the Schwinger effect, modelling it as inducing a charged current $\mathbf{J}$ which can be represented in terms of Ohm's law through electric and magnetic conductivities. These conductivities both backreact on the production of gauge fields and quantify the growth in the energy density of charge carriers over time.

Our work conceptually improves upon this effective model of the Schwinger effect in axion inflation. The first improvement is regarding the ambiguous Ohmic representation of the induced current $\mathbf{J}$, either through the electric picture, $\mathbf{J} = \sigma_E \mathbf{E}$, or the magnetic picture, $\mathbf{J} = \sigma_B \mathbf{B}$. We resolve this ambiguity by relaxing the assumption that axion inflation generates perfectly (anti)collinear electric and magnetic fields. Starting from a generic configuration of the gauge fields in the comoving frame, we boost to the collinear frame where we compute the induced current based on the known result for the Schwinger-induced current in this configuration, Eq. (6.10), and then boost back to the comoving frame. The result is the unambiguous mixed-picture representation, $\mathbf{J} = \sigma_E \mathbf{E} + \sigma_B \mathbf{B}$, given in Eq. (6.29). The second improvement is regarding the backreaction of the conducting medium on gauge-field production. We identified a characteristic physical momentum scale k_S , Eq. (6.50), which reflects the typical separation of charge carriers in the conductive medium based on the strength of the electromagnetic field. Physically, gauge modes with wavelengths below this characteristic length scale should be insensitive to the damping conductive medium. In our improved model, we therefore enforce that damping effects are only affecting modes with wavenumbers $k \leq k_S$.

An important aspect about the changes we made to the existing effective model is that they do not alter the overall structure of the equations of motion of axion inflation. This entails that the new effective model we have devised here is still compatible with the idea of the gradient-expansion formalism (GEF). In particular, the mixed-picture representation of the induced current $\mathbf{J}$ does not structurally differ from the existing electric and magnetic pictures, and is therefore a simple, yet effective way of accounting for Schwinger pair production through non-collinear gauge fields while remaining within the original spirit of the GEF proposed in Ref. [128]. Scale-dependent effects prove to be more challenging to account for, and incorporating the pair-creation scale k_S into the GEF, fundamentally requires us to work with what essentially amounts to a double copy of the usual GEF equations. In this sense, the model with scale-dependent damping below k_h is an appealing compromise, as it does not introduce the technical challenges that arise when incorporating the second scale k_S into the GEF, but it still retains the conceptual idea of modelling Schwinger damping as a scale-dependent effect.

The fact that our revised model is still compatible with the GEF proves to be important, as this allowed us to perform the comprehensive analysis in Sec. 6.4, where we assessed the different modelling assumptions that one may consider within the effective framework we have proposed here. This permitted us to clarify that the changes we have proposed to the effective model are not only conceptually well-motivated, but appear to be robust in the context of previous modelling assumptions. The mixed picture essentially clarifies the modelling uncertainty that arises from the ambiguous choice between the electric and the magnetic picture, while scale-dependent damping effects entail that the abundance of produced fermions or gauge fields during axion inflation may be larger than what had previously been anticipated based on the old model of scale-independent Schwinger damping. Such differences can even induce variations within one order of magnitude, which are ultimately important to take into consideration when assessing the phenomenological

relevance of Schwinger pair production, for instance regarding the Schwinger effect as a reheating channel (as in Ref. [124]), as a means for dark-matter production (as in Refs. [125, 126]), or in view of the imprint of gauge-field production on inflationary perturbations (as in Ref. [362]).

More generally, having a method as the GEF which is able to capture the dynamical backreaction of charge carriers during axion inflation has proven to be highly relevant, as it allowed us to reveal the existence of the new regime of fermion-tempered backreaction. Important in this regard is the fact that this regime appears to persist over a large range of e -folds, and therefore required a technique such as the GEF, which is able to resolve the dynamics of inflation for such a large dynamical range in the first place. This underpins one of the advantages of the GEF we have highlighted in chapter 4.

The knowledge that a regime such as fermion-tempered backreaction exists is itself of significance for two reasons. For one, it clarifies that the Schwinger effect does not necessarily suppress gauge-field backreaction onto the inflaton's dynamics, which had not been explicitly demonstrated prior to the work presented here. Moreover, the persisting oscillatory pattern over multiple e -folds that both the gauge-field energy density and the fermion energy density exhibit during a stage of fermion-tempered backreaction can manifest itself phenomenologically: as we shall explicitly see in chapter 9, these persisting bursts of gauge-field production are then also reflected in the spectrum of gauge-field-sourced gravitational waves from axion inflation.

We should however also note upon the limitations of the GEF approach to treating the Schwinger effect. For one, as highlighted in chapter 4, our effective treatment of Schwinger pair creation implicitly assumes that inflaton inhomogeneities are dynamically irrelevant. The dynamical role of inflaton inhomogeneities in the presence of efficient Schwinger pair production still remains an open question. For a select set of benchmark points the lattice simulations of Ref. [134] indeed confirm that inhomogeneous-backreaction effects are absent if Schwinger damping suppresses gauge-field backreaction effects. This was demonstrated by relying on the same effective treatment of the Schwinger effect presented in this chapter. However, what remains unclear is if this conclusion also holds for the case of fermion-tempered backreaction. Indeed, it is conceivable that the onset of strong-backreaction effects in this dynamical regime could also entail the growth of axion inhomogeneities, which could significantly alter the oscillatory motion of the inflaton–fermion–gauge-field system we have seen in Fig. 6.6. This question should be assessed in future work using lattice techniques. However, in this case, it is likely that input from the GEF is necessary, as the lattice may not be suitable to study an effect which persists for such an extended duration of inflation (cf. Sec. 4.4).

Beyond this, it is also clear that the effective model we have put forth in this chapter only serves to establish a simple picture regarding the backreaction of the Schwinger effect onto gauge-field production, as it inherently assumes an instantaneous response between charge carrier production and backreaction. The validity of this approach has been confirmed by other methods, most notably using a hydrodynamic or kinetic framework [421, 424–427], for as long as slow-roll dynamics persist, i.e., for a slowly-varying inflationary background. However, it is not clear if these conclusions transfer to a stage of fermion-tempered backreaction, where the inflaton's motion is already subjected to significant gauge-field-induced friction.

Despite these limitations, the effective model we have proposed here, and its practical implementation into the GEF, are an important tool for further exploring the dynamical and phenomenological implications of the Schwinger effect in axion inflation. For the purposes of this thesis, the GEF version we have devised here is the backbone of our analysis in chapter 9, where we use it as a basis to determine how Schwinger damping of gauge-field production affects the creation of gravitational waves in axion inflation. Beyond this concrete application within this thesis, the GEF variant we have developed through the course of this chapter can also be applied to the study of other scenarios in which the Schwinger effect is phenomenologically important, e.g., to expand

upon works who study the Schwinger effect as a channel for dark-matter production, reheating, or to assess its importance in the context of baryogenesis. To aid such future work, our Python package, the GEFF [\[7\]](#), provides the necessary numerical framework.

Part III

Primordial perturbations in axion inflation

Chapter 7

Cosmological perturbation theory

In the previous part of this thesis, we have taken an extensive look at the complex dynamical structure that emerges in axion inflation as a consequence of gauge-field production, using the gradient-expansion formalism (GEF) as the tool to resolve these dynamics. In this, we have entirely focussed on how the different matter constituents in this model, the gauge fields, the inflaton, and even charged particles, evolve as part of the homogeneous and isotropic background during inflation. In this last part of this thesis, we now change our perspective, and we instead assess how the continuous creation of matter fields in axion inflation manifests itself in the spectrum of perturbations to this inflationary background, and to assess the possibility for detecting the signatures of these dynamical processes in the future, for instance through gravitational-wave experiments.

Specifically, what we focus on in the next two chapters is how the amplified electromagnetic fields that are created via the evolving inflaton field during axion inflation can themselves source inflationary scalar and tensor perturbations in addition to the conventional mechanism in slow-roll inflation, where inflaton and metric perturbations are only amplified via their interaction with the rapidly changing gravitational background during inflation. The computation of these effects occurs within the established framework of cosmological perturbation theory, which we had already briefly introduced in Sec. 1.4, albeit with a specific focus on slow-roll inflation, i.e., for a model in which only perturbations to the inflaton field and the spacetime metric are relevant. For the next two chapters, we however require a more general version of this formalism which also accounts for the possibility that other matter fields, like gauge fields or charge carriers, are present during inflation. In the following we therefore give a slightly more in-depth presentation of cosmological perturbation theory to establish a common mathematical and notational basis that we can rely on for our subsequent two chapters.

The discussion we present below can be found in various textbooks and reviews on cosmology. Our discussion is oriented along that of Refs. [137, 138, 190–192]. In particular, given that their choice of metric signature matches ours, we have mainly oriented ourselves along the discussions in Refs. [137, 190].

7.1 A closer look at perturbation theory

Fundamentally, cosmological perturbation theory is a formalism which is designed to describe the evolution of small deviations from spatial homogeneity and isotropy on a homogeneous and isotropic background described by the FLRW metric. For a generic energy–momentum tensor $T_{\mu\nu}$, the basic idea is to decompose the Einstein equations,

$$G_{\mu\nu} = \frac{1}{M_{\text{P}}^2} T_{\mu\nu}, \quad (7.1)$$

into two separate sets of equations by means of a perturbative Ansatz for both the metric and the energy–momentum tensor:

$$g_{\mu\nu}(\eta, \mathbf{x}) = \bar{g}_{\mu\nu}(\eta) + \delta g_{\mu\nu}(\eta, \mathbf{x}), \quad (7.2a)$$

$$T_{\mu\nu}(\eta, \mathbf{x}) = \bar{T}_{\mu\nu}(\eta) + \delta T_{\mu\nu}(\eta, \mathbf{x}). \quad (7.2b)$$

The background metric, $\bar{g}_{\mu\nu}$, and energy–momentum tensor, $\bar{T}_{\mu\nu}$, then describe the spatially homogeneous and isotropic FLRW spacetime, while the perturbations, $\delta g_{\mu\nu}$ and $\delta T_{\mu\nu}$ encode deviations from isotropy and homogeneity, which one treats as linear perturbations. Specifically, for the case of axion inflation, the energy–momentum tensor in question describes the inflaton field, the gauge fields, and possibly charged fields, see Eqs. (3.9) and (6.14).

The formalism of cosmological perturbation theory can be understood as a strategic approach to the study of perturbations by making use of the underlying symmetries of the background spacetime, and the invariance of General Relativity under coordinate transformations. In the following, we go through the basic ideas underlying this construction, starting with the separate discussion of metric and matter perturbations before combining them to obtain a set of linearized Einstein equations describing their evolution in time and how they source each other.

7.1.1 Metric perturbations

To start, we begin by separating the metric $g_{\mu\nu}$ into a background metric describing the homogeneous and isotropic FLRW spacetime, $\bar{g}_{\mu\nu}$, and a perturbed metric $\delta g_{\mu\nu}$. We describe FLRW space in conformal time η , such that we can write the metric as

$$g_{\mu\nu}(\eta, \mathbf{x}) = \bar{g}_{\mu\nu}(\eta) + \delta g_{\mu\nu}(\eta, \mathbf{x}) = a^2(\eta) (\eta_{\mu\nu} + h_{\mu\nu}(\eta, \mathbf{x})). \quad (7.3)$$

The perturbative Ansatz assumes that $h_{\mu\nu}$ is a small perturbation to the $\eta_{\mu\nu}$, i.e., $|h_{\mu\nu}| \ll 1$. This implies that the inverse metric to linear order is

$$g^{\mu\nu} = \frac{1}{a^2} (\eta^{\mu\nu} - h^{\mu\nu}) \quad (7.4)$$

and indices of perturbed tensors can be lowered and raised by $\eta^{\mu\nu}$.

The metric perturbation $h_{\mu\nu}$ is symmetric, and thus, in total describes 10 degrees of freedom. These are conveniently organized by applying the *scalar–vector–tensor decomposition* (SVT decomposition), which we have already briefly touched upon in Sec. 1.4. From a mathematical perspective, this decomposition entails writing $h_{\mu\nu}$ as

$$h_{00} = 2A, \quad (7.5a)$$

$$h_{0i} = \partial_i B + B_i, \quad (7.5b)$$

$$h_{ij} = 2C\delta_{ij} + 2\Delta_{ij}E + 2\partial_{(i}E_{j)} + 2E_{ij} \quad (7.5c)$$

where we have introduced the notation

$$\partial_{(i}E_{j)} = \frac{1}{2} (\partial_i E_j + \partial_j E_i), \quad (7.6)$$

$$\Delta_{ij}E = \left(\partial_i \partial_j - \frac{1}{3} \delta_{ij} \Delta \right) E. \quad (7.7)$$

The variables we have introduced here are a set of four scalar perturbations, A , B , C , and E , two divergenceless vector perturbation, B_i and E_i with $\partial^i B_i = \partial^i E_i = 0$, and a divergenceless,

traceless, and symmetric tensor perturbation, E_{ij} with $\partial^i E_{ij} = 0$ and $\delta^{ij} E_{ij} = 0$.¹ Thus, this decomposition of $h_{\mu\nu}$ is simply a reformulation of its 10 degrees of freedom into 4 scalar, 4 vector, and 2 tensor degrees of freedom, and can be applied on general grounds.

We have already noted the advantage of the SVT decomposition in Sec. 1.4 in that it entails that scalar, vector, and tensor perturbations can be studied separately from one another at linear order in perturbation theory. The reason for this is the underlying symmetry of the background metric $\bar{g}_{\mu\nu}$. FLRW space is fully isotropic and homogeneous, and thus remains invariant under spatial translations and rotations. To understand the SVT decomposition, it is then convenient to consider a perturbation $X_{\mathbf{k}}(\eta)$ in Fourier space with a definite wavevector $\mathbf{k}$. The wavevector $\mathbf{k}$ selects a preferred direction, which reduces the rotational invariance of FLRW space to rotations about $\mathbf{k}$ corresponding to an $SO(2)$ symmetry group. A perturbation can then be classified according to the irreducible representations of this group, which can be labelled by an integer helicity. In particular, this means that a perturbation $X_{\mathbf{k}}$ can be decomposed into helicity scalars (helicity zero), transverse vectors (helicity one, $k^i V_i = 0$), and symmetric transverse–traceless tensors (helicity two, $k^i H_{ij} = H^i_i = 0$). Since the linearized set of equations describing perturbations respect the rotational symmetry of the background, perturbations of different helicity do not mix at linear order. The SVT decomposition which we applied to $h_{\mu\nu}$ is simply the real-space formulation of this helicity decomposition, with derivatives in place of the wavevector components k^i .

Moving on from the SVT decomposition, the other thing we should then consider is the fact that General Relativity is invariant under coordinate transformations. As such, the metric perturbation $h_{\mu\nu}$ is not uniquely defined, since we are always free to perform an infinitesimal coordinate transformation of the kind

$$x^\mu \rightarrow x^\mu + \xi^\mu(\eta, \mathbf{x}). \quad (7.8)$$

Treating ξ^μ as a perturbation, $|\xi^\mu| \ll 1$, this change of coordinates, at linear order in perturbation theory, amounts to the following redefinition of $h_{\mu\nu}$

$$h_{\mu\nu} \rightarrow h_{\mu\nu} - \partial_\mu \xi_\nu - \partial_\nu \xi_\mu - 2\mathcal{H}\xi^0 \eta_{\mu\nu}, \quad (7.9)$$

The freedom to choose the shift vector ξ^μ can be interpreted as a gauge freedom in analogy to, e.g., electrodynamics.

Applying the SVT decomposition to ξ^μ , we can write it as

$$\xi^0 = \alpha, \quad \xi^i = \partial^i \beta + \beta^i, \quad (7.10)$$

such that the gauge freedom amounts to choosing two scalars, α and β , and one divergenceless vector β_i . In terms of the metric perturbations which we defined in Eq. (7.5), a change of coordinates amounts to the redefinition

$$A \rightarrow A - \alpha' - \mathcal{H}\alpha, \quad (7.11a)$$

$$B \rightarrow B + \alpha - \beta', \quad B_i \rightarrow B_i - \beta'_i, \quad (7.11b)$$

$$C \rightarrow C - \mathcal{H}\alpha - \frac{1}{3}\Delta\beta, \quad (7.11c)$$

$$E \rightarrow E - \beta, \quad E_i \rightarrow E_i - \beta_i, \quad E_{ij} \rightarrow E_{ij}, \quad (7.11d)$$

Thus, not all the perturbation variables we have defined in Eq. (7.5) are gauge-invariant. In particular, two scalar and one vector metric perturbations can be removed by choosing α , β and

¹In our convention, these variables are raised and lowered using the Euclidean metric, i.e., $\delta^{ik} E_{kj} = E^i_j$.

β_i . A convenient choice we use in this thesis is *Newtonian gauge*, which corresponds to setting $B = E = B^i = 0$.

To obtain a clearer perspective on the physical metric perturbations, it is then convenient to introduce gauge-invariant perturbation variables based on Eq. (7.11). These are the Bardeen variables

$$\Psi \equiv A + \mathcal{H}(B - E') + (B - E')', \quad (7.12a)$$

$$\Phi \equiv -C - \mathcal{H}(B - E') + \frac{1}{3}\Delta E, \quad \Phi_i \equiv E'_i - B_i, \quad (7.12b)$$

We have already briefly introduced the two scalar Bardeen variables, Φ and Ψ , in Sec. 1.4, noting that they can be understood as generalizations of the Newtonian potential for weak gravity.

Furthermore, we also match the definition of the transverse-traceless tensor perturbation given in chapter 1, and define the variable

$$h_{ij}^{TT} \equiv \frac{1}{2}E_{ij}. \quad (7.13)$$

With this definition, h_{ij}^{TT} can then be interpreted as describing propagating gravitational waves in FLRW space.

7.1.2 Matter perturbations

A general form for the energy–momentum tensor is that of an anisotropic fluid,

$$T^\mu{}_\nu(\eta, \mathbf{x}) = [\rho(\eta, \mathbf{x}) + p(\eta, \mathbf{x})]U^\mu(\eta, \mathbf{x})U_\nu(\eta, \mathbf{x}) - \delta^\mu_\nu p(\eta, \mathbf{x}) - \Sigma^\mu{}_\nu(\eta, \mathbf{x}). \quad (7.14)$$

Here, we have defined the total energy density, ρ , pressure p , and anisotropic stress $\Sigma_{\mu\nu}$, in a rest frame defined by the four-velocity U^μ . We define anisotropic stress to be orthogonal to U^μ , $U^\mu \Sigma_{\mu\nu} = 0$, allowing us to set $\Sigma^0{}_0 = \Sigma^0{}_i = 0$.

These variables can be further decomposed into background variables and perturbations. An isotropic background necessarily has vanishing anisotropic stress, so $\Sigma^i{}_j$ vanishes at the background level. Pressure and energy density are instead separated into a homogeneous background contribution and an inhomogeneous perturbation;

$$\rho(\eta, \mathbf{x}) = \bar{\rho}(\eta) + \delta\rho(\eta, \mathbf{x}), \quad (7.15a)$$

$$p(\eta, \mathbf{x}) = \bar{p}(\eta) + \delta p(\eta, \mathbf{x}), \quad (7.15b)$$

With this separation, the components of the energy–momentum tensor are then given by

$$\bar{T}^0{}_0 = \bar{\rho}, \quad \delta T^0{}_0 = \delta\rho, \quad (7.16a)$$

$$\bar{T}^i{}_0 = 0, \quad \delta T^i{}_0 = S^i, \quad (7.16b)$$

$$\bar{T}^i{}_j = -\bar{p}\delta^i_j, \quad \delta T^i{}_j = -\delta p\delta^i_j - \Sigma^i_j, \quad (7.16c)$$

Here, the quantity S^i can be understood as a perturbation to the 3-momentum density of matter, $S^i = (\bar{\rho} + \bar{p})v^i$ where v^i is the physical spatial velocity of the fluid. Note that our convention here is once again that indices of S^i and Σ_{ij} are raised and lowered using the Euclidean metric.

We can further apply the SVT decomposition to S^i and Σ_{ij} ,

$$S_i = \partial_i q + q_i, \quad (7.17a)$$

$$\Sigma_{ij} = \Delta_{ij}\sigma + \partial_{(i}\sigma_{j)} + \sigma_{ij}^{TT}, \quad (7.17b)$$

such that, again, q_i , σ_i , and σ_{ij}^{TT} are divergenceless, with the latter being also traceless.

Defining the perturbation variables as we have done in Eq. (7.16) is useful when considering the case when the energy–momentum tensor accounts for different matter species X , such that from $T_{\mu\nu} = \sum_X (T_X)_{\mu\nu}$. Our definition implies that perturbations for each species X are additive

$$\bar{\rho} = \sum_X \bar{\rho}_X, \quad \delta\rho = \sum_X \delta\rho_X, \quad \bar{p} = \sum_X \bar{p}_X, \quad \delta p = \sum_X \delta p_X, \quad (7.18a)$$

$$S_i = \sum_X (S_i)_X, \quad \Sigma_{ij} = \sum_X (\Sigma_X)_{ij}. \quad (7.18b)$$

In particular, this is relevant for our application of perturbation theory to the case of axion inflation in the next chapters. As we have seen in previous chapters, the energy–momentum tensor in this model takes contributions from both the inflaton field and gauge fields, and may even include charge carriers if one also accounts for Schwinger pair production.

As for the metric, not all perturbation variables are gauge-invariant. Rather, the following transform non-trivially under the infinitesimal change of coordinates in Eq. (7.8);

$$\delta\rho \rightarrow \delta\rho - \alpha\bar{\rho}', \quad \delta p \rightarrow \delta p - \alpha\bar{p}', \quad (7.19a)$$

$$q \rightarrow q + (\bar{\rho} + \bar{p})\beta', \quad q_i \rightarrow q_i + (\bar{\rho} + \bar{p})\beta'_i. \quad (7.19b)$$

Note that all components of the anisotropic-stress tensor, Σ_{ij} , are gauge invariant.

We can then introduce gauge-invariant combinations of metric and matter perturbations. For our work, we rely on the comoving curvature perturbation variable to describe scalar perturbations, which we have already defined in Eq. (1.74) for describing the scalar power spectrum in the case of slow-roll inflation. Based on the conventions used in this chapter, its general definition for a system with a generic matter constituents is

$$\zeta = -C + \frac{1}{3}\Delta E - \mathcal{H} \left(B + \frac{q}{\bar{\rho} + \bar{p}} \right) \quad (7.20)$$

It is a straight-forward exercise to determine that this general expression for ζ reduces to Eq. (1.74) for the case of Newtonian gauge, $B = E = 0$ and for a perturbed energy–momentum tensor of a scalar field.

7.1.3 The linearized Einstein equations

With the separation of $g_{\mu\nu}$ and $T_{\mu\nu}$ into background and perturbation in Eq. (7.2), Einstein equations can now formally be separated into two equations; one background equation

$$\bar{G}^\mu{}_\nu = \frac{1}{M_{\text{P}}^2} \bar{T}^\mu{}_\nu, \quad (7.21)$$

and a perturbed equation

$$G^\mu{}_\nu - \bar{G}^\mu{}_\nu = \frac{1}{M_{\text{P}}^2} \delta T^\mu{}_\nu, \quad (7.22)$$

where $\bar{G}^\mu{}_\nu$ depends purely on the background metric $\bar{g}_{\mu\nu}$. For completeness, these can be accompanied by the continuity equation for the energy–momentum tensor, $\nabla_\mu T^\mu{}_\nu = 0$, which can be similarly decomposed into background and perturbed equations.

The background equation simply yields the usual Friedmann equations;

$$\mathcal{H}^2 = \frac{a^2 \bar{\rho}}{3M_{\text{P}}^2}, \quad 2\mathcal{H}' + \mathcal{H}^2 = -\frac{a^2 \bar{p}}{M_{\text{P}}^2}, \quad (7.23)$$

with the familiar continuity equation

$$\bar{\rho}' + 3\mathcal{H}(\bar{\rho} + \bar{p}) = 0. \quad (7.24)$$

Going beyond the background, the perturbed Einstein equations can be expanded to linear order in the perturbation variables for both the metric and the energy–momentum tensor as we have introduced them in the previous two sections. Here is where the power of the SVT decomposition manifests itself, as it ensures that the equations for scalar, vector, and tensor perturbations are merely coupled via the background Friedmann equation. We can therefore discuss the equations for each of these three sectors separately from the others.

Scalar perturbations: The scalar sector is described by four metric perturbations (A, B, C, E) and four matter perturbations ($\delta\rho, \delta p, q, \sigma$), with a residual gauge freedom to choose two scalar functions (α, β). We use the gauge freedom to work in Newtonian gauge, setting $B = E = 0$. The remaining two metric perturbations can then be conveniently expressed in terms of the Bardeen variables, Ψ and Φ , which, in Newtonian gauge, are simply given by

$$\Psi = A, \quad \Phi = -C. \quad (7.25)$$

The linearized Einstein equations for these perturbations are then

$$2M_{\text{P}}^2 \Delta \Phi = a^2 (\delta\rho - 3\mathcal{H}q), \quad (7.26a)$$

$$2M_{\text{P}}^2 [\Phi'' + \mathcal{H}(\Psi' + 2\Phi')] + (2\mathcal{H}' + \mathcal{H}^2)\Psi = a^2 \delta p, \quad (7.26b)$$

$$2M_{\text{P}}^2 (\Phi' + \mathcal{H}\Psi) = -a^2 q, \quad (7.26c)$$

$$M_{\text{P}}^2 (\Phi - \Psi) = a^2 \sigma. \quad (7.26d)$$

It may be noted that the last equation implies $\Phi = \Psi$ in the absence of anisotropic stress, $\sigma = 0$, which can further simplify computations. The other equation to take note of is Eq. (7.26a): in the absence of momentum perturbations, $q = 0$, this is simply the usual equation for a Newtonian potential. This underpins our previous statement that the Bardeen variables, Φ and Ψ may be understood as generalizations of the Newtonian potential in weak gravity. It also explains the name of the Newtonian gauge, as it evidences this connection.

In particular, we should take note in the above equations that any scalar perturbation to the matter sector in a given model can source metric perturbations. In the next chapter, we shall build on this idea to investigate how gauge fields can source metric perturbations directly. To this end, it is useful to work in terms of gauge-invariant variables, namely in terms of the curvature perturbation variable ζ . In Newtonian gauge, it takes the following form

$$\zeta = \Phi - \mathcal{H} \frac{q}{\bar{\rho} + \bar{p}}, \quad (7.27)$$

which directly establishes a connection to the Bardeen potential Φ . Using Eq (7.23) and Eq (7.26d), ζ may then also be written in the following gauge-invariant way;

$$\zeta = \Phi + \frac{\mathcal{H}}{\mathcal{H}^2 - \mathcal{H}'} (\Phi' + \mathcal{H}\Psi). \quad (7.28)$$

In the next chapter, we shall use this form for ζ as a basis to establish a master equation for the evolution of ζ which accounts for the combined evolution of scalar metric, inflaton, and gauge-field perturbations in axion inflation.

Vector perturbations: This sector is composed of two metric perturbation (B_i, E_i) and two matter perturbations (q_i, σ_i) with one residual gauge freedom (β_i). We use this gauge-freedom to set $B_i = 0$. In this gauge, it is once again simplest to express the linearized Einstein equations through the Bardeen variable Φ_i , which is simply $\Phi_i = E'_i$.

The dynamics of vector perturbations are then described by

$$\Delta\Phi_i = -\frac{2a^2}{M_{\text{P}}^2}q_i, \quad \Phi'_i + 2\mathcal{H}\Phi_i = \frac{\sigma_i}{M_{\text{P}}^2}, \quad (7.29)$$

with the only non-trivial equation following from covariant energy conservation being

$$q'_i + 4\mathcal{H}q_i + \frac{1}{2}\Delta\sigma_i = 0 \quad (7.30)$$

An important thing to note about vector perturbations is that, for vanishing anisotropic stress, $\sigma_i = 0$, Eq. (7.29) merely admits a solution which is decaying with the scale factor. As such, vector perturbations are typically considered to be irrelevant for cosmology. Even if they are sourced by anisotropic stress at some point in the early Universe, as soon as this source becomes inactive, the perturbations die away.

It should however be noted that vector perturbations can be sourced by gauge-field perturbations during inflation. At higher order in perturbation theory, these could source other, relevant perturbations, for instance tensor perturbations, which are linked to GW production, see for instance Ref. [428]. To our knowledge, this idea has, however, not been considered in the context of axion inflation, and we shall also not explore it further in this thesis.

Tensor perturbations: Finally, the description of the remaining tensor perturbations is remarkably simple. There is a single tensor metric perturbation (h_{ij}^{TT}) composed of two degrees of freedom, and a single matter perturbation (σ_{ij}^{TT}) with no further residual gauge freedom to remove either of them. These two variables are directly related by a single non-trivial equation:

$$(h_{ij}^{\text{TT}})'' + 2\mathcal{H}(h_{ij}^{\text{TT}})' - \Delta h_{ij}^{\text{TT}} = \frac{2a^2}{M_{\text{P}}^2}\sigma_{ij}^{\text{TT}}. \quad (7.31)$$

Thus, we can see the well-known result that tensor metric perturbations, i.e., gravitational waves, are sourced by non-vanishing anisotropic stress. In chapter 9, we consider how the anisotropic stress from gauge-field production sources gravitational waves, and we also estimate how fermions sourced via the Schwinger effect could add to the production of gravitational waves.

With this, we have established all that we require for the further exploration of inflationary perturbations in the context of axion inflation in the next chapters. Without further ado, our first analysis is then regarding the sourcing of scalar perturbations from inflationary gauge fields.

Chapter 8

Gauge-field-induced scalar perturbations

This chapter is based on Ref. [2], except for Sec. 8.3 which contains original results that have been developed specifically for this thesis. The detailed contributions of the author to Ref. [2] are outlined at the beginning of this thesis.

In the previous chapter, we have established the general formalism of cosmological perturbation theory that describes how small deviations from a perfectly isotropic and homogeneous Universe are sourced and evolve over time. In this chapter, we now apply this formalism to the study of scalar perturbations in axion inflation, focussing on the simpler version of this model, in which the Schwinger effect is assumed to be ineffective.

In this model, there are two complementary modes for the creation of scalar perturbations [55–57]. The first is the conventional inflationary mechanism we have discussed in chapter 1; the rapid expansion of the Universe amplifies vacuum fluctuations in the inflaton field and the metric, which then contribute to the spectrum of scalar perturbations. The second mechanism relates to the enhanced background of gauge fields that are sourced by the rolling inflaton field during axion inflation. This gauge-field background itself sources scalar perturbations, which we shall refer to as *gauge-field-induced scalar perturbations*.

As we have introduced in chapter 3, gauge-field-induced scalar perturbations are a central prediction of axion inflation with important phenomenological implications. Unlike the conventional spectrum of vacuum scalar perturbations, which are generically predicted by slow-roll inflation to be scale-invariant, and predominantly Gaussian (cf. chapter 1), the scalar perturbations sourced by the gauge-field background exhibit deviations from a Gaussian spectrum [56, 57, 65]. Such non-Gaussianities are strongly constrained by CMB measurements [15, 49] and therefore limit the efficiency of gauge-field production in the early phases of axion inflation [56, 57]. Moreover, a continuous amplification of the gauge-field background can lead to the enhancement of scalar perturbations at late times during axion inflation. Scalar perturbations sourced during this stage of inflation can affect the distribution of matter in the radiation-dominated era, leading to the formation of overdense regions of spacetime that can collapse gravitationally to form primordial black holes (PBHs) (see e.g., Ref. [253] for a review). This inflationary production channel for PBHs can lead to an overabundance of PBH dark matter, which, then again, constraints the efficiency of gauge-field production during axion inflation [84, 89, 90, 107].

Previous studies of scalar perturbations from axion inflation focussed entirely on gauge fields as a source of inflaton perturbations (see for instance Ref. [57] for a detailed discussion of this calculation). However, as we have already noted upon in the last chapter, gauge fields may also generically induce metric perturbations via their gravitational effect, which previous calculations in the literature did not account for. In the first part of this chapter, we therefore devise a formalism in which both contributions from gauge-field-induced inflaton and metric perturbations are consistently taken into account. Moreover, the framework we establish here can also be used to

assess how the amplified electromagnetic background radiation that is consistently present during axion inflation backreacts on the evolution of scalar perturbations itself. To our knowledge, this effect has thus far not been considered in the literature on axion inflation.

Once the general framework has been established, we then consider its concrete application in an analytical and a numerical case study. Our analytical analysis is based on the assumption that gauge-field backreaction is entirely irrelevant, both for the background dynamics, and regarding perturbations. This analytical study permits a first controlled assessment regarding the relevance of gauge-field-sourced metric perturbations, and we perform a comprehensive comparison to existing analytical results. Moreover, we also apply these results to reevaluate the existing CMB bound on scalar non-Gaussianity in axion inflation.

Expanding upon this first analysis, our numerical case study instead is meant to assess the relevance of sourced metric perturbations in the regime of strong gauge-field backreaction, where our prior analytical calculation is no longer applicable. Here, we rely on the gradient-expansion formalism (GEF) to determine the dynamical evolution of the inflationary background, and then compute the scalar power spectrum based on this knowledge. We apply this technique in a staggered approach, first assuming that gauge-field backreaction only affects the inflationary background, but not the evolution of scalar perturbations itself, before finally including backreaction effects for both background and perturbations.

Our numerical approach for studying scalar perturbations based on the GEF complements other numerical frameworks for computing the scalar power spectrum from axion inflation which exist in the literature, i.e., the approximate numerical calculations based on the mode-by-mode approach in Refs. [107, 109], and direct extractions of the scalar power spectrum from lattice simulations [380, 381]. We use this method in this chapter to assess the relevance of gauge-field-induced metric perturbations in the strong-backreaction regime, which prior analyses did not account for, but it also has applications beyond this use. A numerical method for computing scalar perturbations is, for instance, also relevant for the study of PBH formation in axion inflation. The necessary amplification of scalar perturbations needed for PBH formation to be effective is likely only realized in coexistence with strong gauge-field backreaction effects [109], and thus requires numerical tools to assess these possibilities.

The discussion in this chapter is structured as follows: In Sec. 8.1, we present the derivation of our novel formalism to study both the electromagnetic creation of metric and inflaton perturbations. In Sec. 8.2, we apply this formalism in an analytical case study to axion inflation in the weak-backreaction regime. Section 8.3 is dedicated to our numerical study of scalar perturbations. Finally, in Sec. 8.4, we summarize our results and provide an outlook regarding future applications of our formalism, and research topics that warrant further investigation.

8.1 The formalism

The formalism we present in this section builds on a consistent treatment of gauge-field perturbations within the framework of cosmological perturbation theory presented in the previous chapter. In doing so, we assume that inflationary gauge fields source perturbations in scalar curvature, while neglecting the backreaction of metric perturbations on the evolution of the gauge-field source. At the core of this formalism is the gauge-invariant treatment of cosmological perturbations in the presence of inhomogeneous seeds developed in Refs. [429, 430]. The main difference between our formalism and the one in Refs. [429, 430], is that, in axion inflation, the gauge-field sources also interact via a non-gravitational interaction with the inflaton, and can backreact on the inflationary evolution. The backreaction of the gauge field on the inflationary evolution can also affect the evolution of scalar perturbations, which our derivation consistently takes into account.

In the first part of our discussion, we derive a set of equations describing the evolution of scalar perturbations that account for (i) linear perturbations of the inflaton field, (ii) scalar linear perturbations of the spacetime metric, and (iii) first-order perturbations to the energy–momentum tensor of the gauge field. Based on these equations, we then derive a master equation for the curvature perturbation variable ζ , which accounts for all these sources of scalar perturbations in axion inflation. We then present a general strategy for solving this master equation for ζ , and we derive general expressions for the two- and three-point correlation function of ζ in Fourier space; the scalar power spectrum and the bispectrum. These general expressions are then the basis for our subsequent applications of our formalism.

It should be noted that, for the purposes of this thesis, we present only the application of this formalism in the context of axion inflation. In our original work, Ref. [2], where we proposed this formalism for the first time, we studied a more generic inflationary model featuring an inflaton field ϕ , coupled to the electromagnetic field-strength tensor $F_{\mu\nu}$. This model features two generic kinds of interactions, a kinetic coupling $\propto I_1(\phi)F_{\mu\nu}F^{\mu\nu}$, and an axial coupling $\propto I_2(\phi)F_{\mu\nu}\tilde{F}^{\mu\nu}$, with $I_1(\phi)$ and $I_2(\phi)$ functions of the inflaton field. The standard axion inflation model which we consider in this thesis is retrieved by setting $I_1(\phi) = 1$, and $I_2(\phi) = (\beta/M_P)\phi$. The case of a non-trivial kinetic coupling is, for instance, relevant for the study of scalar perturbations in the Ratra model [393], see Refs. [431, 432] for seminal works on this subject.

8.1.1 Scalar perturbations in axion inflation

We begin with the standard action for the axion inflation describing a real pseudoscalar field ϕ coupled to a gauge field A_μ , which reads

$$S = \int d^4x \sqrt{-g} \left[-\frac{M_P^2}{2} R + \frac{1}{2} g^{\mu\nu} \partial_\mu \phi \partial_\nu \phi - V(\phi) - \frac{1}{4} (F_{\mu\nu} F^{\mu\nu}) - \frac{\beta \phi}{4 M_P} (F_{\mu\nu} \tilde{F}^{\mu\nu}) \right]. \quad (8.1)$$

As per our standard notation, $F_{\mu\nu}$ is the field-strength tensor, $\tilde{F}^{\mu\nu}$ its dual, $V(\phi)$ gives the inflaton potential, and β is a dimensionless parameter describing the strength of the axion–vector coupling. For later convenience, we also introduced the notation $(\cdots)$ for the symmetrized product of two operators,

$$(PQ) \equiv \frac{1}{2}(PQ + QP),$$

which may no longer commute after quantization. Although not strictly necessary at this point, we already introduce this symmetrization here on the basis of the correspondence principle for the analogous action written in terms of quantum fields. In previous chapters, we have neglected to include this symmetrization, but we shall need to apply it more carefully in the study of scalar perturbations.¹

From this action one obtains the standard energy–momentum tensor for axion inflation

$$T_{\mu\nu} = \partial_\mu \phi \partial_\nu \phi - g^{\alpha\beta} (F_{\mu\alpha} F_{\nu\beta}) - g_{\mu\nu} \left[\frac{1}{2} \partial_\alpha \phi \partial^\alpha \phi - V(\phi) - \frac{1}{4} (F_{\alpha\beta} F^{\alpha\beta}) \right], \quad (8.2)$$

¹In particular, in chapter 3, this symmetrization could only affect the terms $\langle \mathbf{E} \cdot \mathbf{B} \rangle$, $\langle \mathbf{E}^2 \rangle$, and $\langle \mathbf{B}^2 \rangle$, with $\mathbf{E}$ and $\mathbf{B}$ the electromagnetic three-vector fields. Since the commutators $[E_i, E_i]$, $[B_i, B_i]$ and $[E_i, B_i]$ vanish, this symmetrization was not necessary there, and only appeared when we discussed the GEF variable $\mathcal{G}^{(n)} \propto \langle \mathbf{E} \cdot \text{rot}^n \mathbf{B} + \mathbf{B} \cdot \text{rot}^n \mathbf{E} \rangle$. In this chapter, similar combinations of the electromagnetic operators are relevant for the study of scalar perturbations.

which dictates the dynamical evolution of the metric $g_{\mu\nu}$ via the Einstein equations. Meanwhile, the fields ϕ and $F_{\mu\nu}$ evolve according to

$$\frac{1}{\sqrt{-g}}\partial_\mu [\sqrt{-g}g^{\mu\nu}\partial_\nu\phi] + V_{,\phi}(\phi) = -\frac{\beta}{4M_{\text{P}}} (F_{\mu\nu}\tilde{F}^{\mu\nu}), \quad (8.3a)$$

$$\frac{1}{\sqrt{-g}}\partial_\mu [\sqrt{-g}F^{\mu\nu}] = -\frac{\beta}{M_{\text{P}}}\partial_\mu\phi\tilde{F}^{\mu\nu}, \quad (8.3b)$$

$$\frac{1}{\sqrt{-g}}\partial_\mu [\sqrt{-g}\tilde{F}^{\mu\nu}] = 0. \quad (8.3c)$$

These are simply the standard equations for axion inflation given in Eq. (3.8) with the additional symmetrization for the gauge-field friction term on the right-hand side of Eq. (8.3a).

We now study this system for the case of a perturbed FLRW metric. We shall consider only linear perturbations to the FLRW spacetime such that scalar perturbations can be studied separately from vector and tensor perturbations. The perturbed metric we consider is therefore

$$g_{\mu\nu}dx^\mu dx^\nu = a^2(\eta) [(1 + 2\Psi(\eta, \mathbf{x})) d\eta^2 - (1 - 2\Phi(\eta, \mathbf{x})) \delta_{ij}dx^i dx^j]. \quad (8.4)$$

Here, we have expressed the metric in conformal time η , a is the scale factor, and we have chosen to work in Newtonian gauge such that all scalar metric perturbations can be entirely expressed in terms of the Bardeen variables $\Psi(\eta, \mathbf{x})$ and $\Phi(\eta, \mathbf{x})$ introduced in the previous chapter.

For this metric, we can define electric- and magnetic-field operators in the same way as on the pure FLRW background, i.e., we define them as

$$F_{0i} = a^2 E^i, \quad F_{ij} = -a^2 \varepsilon_{ijk} B^k. \quad (8.5)$$

Here, $\mathbf{E}$ and $\mathbf{B}$ are again the electric and magnetic fields that a comoving observer in the background FLRW spacetime would measure.

The scenario we implicitly assume in the following is one of an inflationary phase driven by the inflaton field. As in previous chapters, we treat all fields ϕ , $\mathbf{E}$, $\mathbf{B}$, Φ , Ψ etc. as quantum fields, whose modes are sub-horizon at early times, $|k\eta| \gg 1$, and which are then excited out of their vacuum state at later times due to the rapid expansion of the Universe (and the inflaton–vector coupling in the case of the gauge field).

Let us then consider the separation of the matter fields ϕ , $\mathbf{E}$ and $\mathbf{B}$ into background and perturbations. For the inflaton field, we apply our usual separation into a homogeneous background inflaton field $\varphi(\eta)$, and a perturbation $\delta\phi(\eta, \mathbf{x})$,

$$\varphi(\eta) \equiv \langle \phi(\eta, \mathbf{x}) \rangle, \quad \delta\phi(\eta, \mathbf{x}) \equiv \phi(\eta, \mathbf{x}) - \varphi(\eta). \quad (8.6)$$

For the gauge field, we have seen in previous chapters that the homogeneous and isotropic background can only be affected by the vacuum expectation values of quadratic gauge-field operators, e.g., $\langle \mathbf{E}^2 \rangle$, $\langle \mathbf{B}^2 \rangle$, $\langle \mathbf{E} \cdot \mathbf{B} \rangle$, $\langle (\mathbf{E} \cdot \text{rot } \mathbf{E}) \rangle$, $\langle (\mathbf{E} \cdot \text{rot}^2 \mathbf{B}) \rangle$, etc.² We therefore describe perturbations to these homogeneous background quantities in terms of the variables

$$\delta_{P_i Q_j} \equiv (P_i Q_j) - \frac{1}{3} \delta_{ij} \langle (\mathbf{P} \cdot \mathbf{Q}) \rangle, \quad (8.7)$$

where $\mathbf{P}$ and $\mathbf{Q}$ can be any spatial vector out of the set $\{\mathbf{E}, \mathbf{B}, \text{rot } \mathbf{E}, \text{rot } \mathbf{B}, \text{rot}^2 \mathbf{E}, \dots\}$. Once again, $\langle \cdot \rangle$ denotes the expectation value during inflation which is independent of position since we

²We do not write the symmetrization $(\cdot)$ for the combination $\mathbf{E}^2$, $\mathbf{B}^2$ and $\mathbf{E} \cdot \mathbf{B}$, as the operators $\mathbf{E}$ and $\mathbf{B}$ commute.

assume a homogeneous and isotropic background.

The dynamics of the background are then described by the usual set of equations. The background expansion is given by the Friedman equations

$$\mathcal{H}^2 = \frac{1}{3M_{\text{P}}^2} \left\{ \frac{\varphi'^2}{2} + a^2 V(\varphi) + \frac{a^2}{2} [\langle \mathbf{E}^2 \rangle + \langle \mathbf{B}^2 \rangle] \right\}, \quad (8.8)$$

$$2\mathcal{H}' + \mathcal{H}^2 = -\frac{1}{M_{\text{P}}^2} \left\{ \frac{\varphi'^2}{2} - a^2 V(\varphi) + \frac{a^2}{6} [\langle \mathbf{E}^2 \rangle + \langle \mathbf{B}^2 \rangle] \right\}, \quad (8.9)$$

where $\mathcal{H} = a'/a$ denotes the comoving Hubble rate. Meanwhile, the inflaton field evolves according to the Klein–Gordon equation,

$$\varphi'' + 2\mathcal{H}\varphi' + a^2 V_{,\phi}(\varphi) = a^2 \frac{\beta}{M_{\text{P}}} \langle \mathbf{E} \cdot \mathbf{B} \rangle, \quad (8.10)$$

and the evolution of the gauge-field expectation values can be cast in the form of the GEF equations as described in chapter 4.

Let us then focus on scalar perturbations. We start by expanding the energy–momentum tensor for our system in Eq. (8.2) to linear order in the perturbation variables Φ , Ψ , $\delta\phi$ and the gauge-field perturbations. Following our definitions in the previous chapter, the scalar matter perturbations to energy density, $\delta\rho$, pressure, δp , momentum density, q , and anisotropic stress σ are then given by

$$\delta\rho = \frac{\varphi'}{a^2} \delta\phi' + V_{,\phi}(\varphi) \delta\phi - \left[\frac{\varphi'^2}{a^2} + \langle \mathbf{E}^2 \rangle \right] \Psi + \left[\langle \mathbf{E}^2 \rangle + 2\langle \mathbf{B}^2 \rangle \right] \Phi + \frac{2M_{\text{P}}^2}{a^2} F_\rho, \quad (8.11a)$$

$$\delta p = \frac{\varphi'}{a^2} \delta\phi' - V_{,\phi}(\varphi) \delta\phi - \left[\frac{\varphi'^2}{a^2} + \frac{1}{3} \langle \mathbf{E}^2 \rangle \right] \Psi + \frac{1}{3} \left[\langle \mathbf{E}^2 \rangle + 2\langle \mathbf{B}^2 \rangle \right] \Phi + \frac{2M_{\text{P}}^2}{a^2} F_p, \quad (8.11b)$$

$$q = -\frac{\varphi'}{a^2} \delta\phi - \frac{2M_{\text{P}}^2}{a^2} F_q, \quad \sigma = \frac{2M_{\text{P}}^2}{a^2} F_\sigma, \quad (8.11c)$$

where, for later convenience, we have introduced the variables F_ρ , F_p , F_q and F_σ ;

$$F_\rho = 3F_p = \frac{a^2}{4M_{\text{P}}^2} [\delta \mathbf{E}^2 + \delta \mathbf{B}^2], \quad F_q = -\frac{a^2}{2M_{\text{P}}^2} \frac{\partial_i}{\Delta} \varepsilon_{ijk} \delta E_j B_k, \quad (8.12a)$$

$$F_\sigma = \frac{a^2}{4M_{\text{P}}^2} \frac{1}{\Delta} [\delta \mathbf{E}^2 + \delta \mathbf{B}^2] - \frac{3a^2}{4M_{\text{P}}^2} \frac{\partial_i \partial_j}{\Delta^2} [\delta E_i E_j + \delta B_i B_j]. \quad (8.12b)$$

These simply encompass scalar perturbations to the electromagnetic energy–momentum tensor rescaled by a convenient factor $a^2/(2M_{\text{P}}^2)$. The inverse Laplacian Δ are easily interpreted in Fourier space as inverse powers of the wavenumber k^2 . They appear because we project out the scalar contributions to $T_{\mu\nu}$.

With these definitions in place, the linearized Einstein equations for the scalar sector in Eq. (7.26) can then be written as

$$\Delta\Phi = \frac{a^2}{2M_{\text{P}}^2} \left\{ \frac{\varphi'}{a^2} (\delta\phi' + 3\mathcal{H}\delta\phi) + V_{,\phi}(\varphi)\delta\phi - \frac{\varphi'^2}{a^2} \Psi \right. \\ \left. + \left[(\Phi - \Psi)\langle \mathbf{E}^2 \rangle + 2\Phi\langle \mathbf{B}^2 \rangle \right] \right\} + F_\rho + 3\mathcal{H}F_q. \quad (8.13a)$$

$$\Phi' + \mathcal{H}\Psi = \frac{\varphi'}{2M_{\text{P}}^2} \delta\phi + F_q, \quad (8.13b)$$

$$\Phi'' + \mathcal{H}(\Psi' + 2\Phi') + (2\mathcal{H}' + \mathcal{H}^2)\Psi = \frac{a^2}{2M_{\text{P}}^2} \left\{ \frac{\varphi'}{a^2} \delta\phi' - V_{,\phi}(\varphi)\delta\phi - \frac{\varphi'^2}{a^2} \Psi \right. \\ \left. + \frac{1}{3} \left[(\Phi - \Psi)\langle \mathbf{E}^2 \rangle + 2\Phi\langle \mathbf{B}^2 \rangle \right] \right\} + \frac{1}{3} (F_\rho + 2\Delta F_\sigma) \quad (8.13c)$$

$$\Phi - \Psi = 2F_\sigma. \quad (8.13d)$$

We can see the gauge field enters here in two ways; first the background terms $\langle \mathbf{E}^2 \rangle$ and $\langle \mathbf{B}^2 \rangle$ couple to both metric and inflaton perturbations, second the energy–momentum tensor generates metric perturbations through F_ρ , F_q and F_σ (F_p being directly proportional to F_ρ , see Eq. (8.12a)). The first kind of interaction can be interpreted as the gauge-field background backreacting on the evolution of scalar perturbations, while the second interaction directly sources metric perturbations.

Starting from Eq. (8.13), we now seek to construct a differential equation for the curvature perturbation ζ introduced in Eq. (7.27). This equation will not only depend on the source terms F_α , but also their time derivatives. Expressions for these can be found starting from the general equations of motion for the gauge field, Eq. (8.3b) and Eq. (8.3c), and expanding to linear order in perturbation theory to determine perturbed Maxwell's equations for $\mathbf{E}$ and $\mathbf{B}$. One ultimately finds

$$F'_\rho = -2\mathcal{H}F_\rho + \Delta F_q - \frac{\beta}{M_{\text{P}}} \frac{a^2 \varphi'}{2M_{\text{P}}^2} \delta_{\mathbf{E} \cdot \mathbf{B}} \quad (8.14a)$$

$$+ \frac{a^2}{2M_{\text{P}}^2} \left\{ \langle \mathbf{E}^2 \rangle (\Phi + \Psi)' - \frac{\beta}{M_{\text{P}}} \langle (\mathbf{E} \cdot \mathbf{B}) \rangle [\delta\phi' + \varphi'(\Phi + \Psi)] - 2\langle (\mathbf{E} \cdot \text{rot } \mathbf{B}) \rangle (\Phi + \Psi) \right\},$$

$$F'_q = -2\mathcal{H}F_q + \frac{2}{3}\Delta F_\sigma + \frac{1}{3}F_\rho \quad (8.14b)$$

$$+ \frac{a^2}{2M_{\text{P}}^2} \left\{ \frac{1}{3} \left[\langle \mathbf{E}^2 \rangle + 2\langle \mathbf{B}^2 \rangle \right] (\Phi + \Psi) - \frac{\beta}{M_{\text{P}}} \langle (\mathbf{E} \cdot \mathbf{B}) \rangle \delta\varphi \right\},$$

$$F'_\sigma = -2\mathcal{H}F_\sigma - \frac{a^2}{2M_{\text{P}}^2} \left(\frac{\delta_{ij}}{\Delta} - 3 \frac{\partial_i \partial_j}{\Delta^2} \right) \left[\frac{\beta}{M_{\text{P}}} \varphi' \delta_{E_i B_j} + \delta_{(\text{rot } \mathbf{E})_i B_j} - \delta_{(\text{rot } \mathbf{B})_i E_j} \right]. \quad (8.14c)$$

For completeness, we can also write down a perturbed Klein–Gordon equation. This can be either directly obtained starting from Eq. (8.3a) or by combining the two Einstein equations Eqs. (8.13a) and (8.13b). It is given by

$$\delta\phi'' + 2\mathcal{H}\delta\phi' - \Delta\delta\phi + a^2 V_{,\phi\phi}(\varphi)\delta\phi = \varphi'(\Psi' + 3\Phi') + 2(\varphi'' + 2\mathcal{H}\varphi')\Psi \\ + a^2 \frac{\beta}{M_{\text{P}}} \left[\langle (\mathbf{E} \cdot \mathbf{B}) \rangle (3\Phi - \Psi) + \delta_{\mathbf{E} \cdot \mathbf{B}} \right]. \quad (8.15)$$

Using these equations, one can now construct a second-order differential equation for the

curvature perturbation variable ζ . For the model we are considering here, which includes metric, inflaton, and gauge-field perturbations, ζ is given by,

$$\zeta = \Phi + \frac{\mathcal{H}}{\mathcal{H}^2 - \mathcal{H}'} \left(\frac{\varphi'}{2M_{\text{P}}^2} \delta\phi + F_q \right), \quad (8.16)$$

which directly follows from Eq. (7.27) and Eq. (8.11).

The full derivation of the differential equation for ζ is presented in App. D. Here, we merely summarize the procedure. First, we use Eqs. (8.13d) and (8.13a) to express $\delta\phi'$ through Φ , $\delta\phi$ and the source functions F_q , F_ρ , and F_σ . Similarly, we determine Φ' as a function of Φ , $\delta\phi$ and the sources F_α using Eqs. (8.13b) and (8.13d). By means of Eqs. (8.13c) and (8.13d) we also write $(\Phi' + \mathcal{H}\Psi)'$ in terms of Φ , $\delta\varphi$, and the sources. The same can be achieved for ζ' by taking the time derivative of Eq. (8.16), which is then also given in terms of Φ , $\delta\varphi$, and the F_α 's. This way, we obtain two linear expressions, $\zeta = \zeta(\Phi, \delta\varphi, F_\alpha)$ and $\zeta' = \zeta'(\Phi, \delta\varphi, F_\alpha)$. These can be inverted to yield expressions $\Phi = \Phi(\zeta, \zeta', F_\alpha)$ and $\delta\varphi = \delta\varphi(\zeta, \zeta', F_\alpha)$. Taking yet another time derivative of ζ' , we then again express it through Φ , $\delta\varphi$, the sources, F_α , and their first time derivatives. Finally, we can use the equations of motions for the F_α 's in Eq. (8.14) to determine ζ'' as a function of ζ' , ζ , the sources, and F'_σ . This yields the desired second-order differential equation for ζ . We perform this derivation in Fourier space, where the final equation takes the form

$$\zeta''_{\mathbf{k}} + p_k \zeta'_k + q_k \zeta_k = \mathcal{S}_k, \quad (8.17)$$

where $\zeta_{\mathbf{k}}$ is the Fourier transform of ζ ,

$$\zeta_{\mathbf{k}}(\eta) \equiv \int \frac{d^3\mathbf{x}}{(2\pi)^{3/2}} \zeta(\eta, \mathbf{x}) e^{-i\mathbf{x}\cdot\mathbf{k}}, \quad (8.18)$$

and the source function $\mathcal{S}_k$ is given by

$$\mathcal{S}_k \equiv r^{(q)} F_q + r^{(\rho)} F_\rho + r_1^{(\sigma)} F_\sigma + r_2^{(\sigma)} F'_\sigma + r^{(\mathbf{E}\cdot\mathbf{B})} \delta_{\mathbf{E}\cdot\mathbf{B}}. \quad (8.19)$$

The coefficients p_k , q_k , and $r^{(\alpha)}$ are functions of the wavevector $k = |\mathbf{k}|$ and of time. They are entirely dependent on the background evolution, including the expectation values $\langle \mathbf{E}^2 \rangle$, $\langle \mathbf{B}^2 \rangle$, and $\langle (\mathbf{E} \cdot \mathbf{B}) \rangle$. The full expressions for these coefficients are lengthy, which is why we only include them in App. D. It should also be noted that the decomposition of the right-hand-side source term in Eq. (8.17) into five contributions is done purely for convenience and holds no deeper physical or mathematical meaning.³

In the case where gauge-field backreaction effects are absent, which corresponds to setting $\langle \mathbf{E}^2 \rangle = \langle \mathbf{B}^2 \rangle = \langle (\mathbf{E} \cdot \mathbf{B}) \rangle = 0$, the coefficients p_k and q_k reduce to

$$p_k = 2 \frac{z'}{z}, \quad q_k = k^2, \quad \text{for } z = \frac{a\varphi}{\mathcal{H}}, \quad (8.20)$$

such that the homogeneous part of Eq. (8.17) simply yields the same differential equation for ζ as in the case of standard slow-roll inflation, see Eq. (1.75). Meanwhile, the right-hand-side source terms in Eq. (8.17) are still present in this scenario, as gauge fields can induce scalar perturbations via their gravitational interaction and via the inflaton–gauge-field interaction in Eq. (8.1) even if they do not directly impact the background dynamics. In particular, for the case $\beta = 0$, i.e., no

³Note also, that in Ref. [2], we decompose the source term into six terms, not five, as we have given it here. The sixth term is proportional to $\delta_{\mathbf{E}^2}$, but its prefactor vanishes exactly for the particular case of axion inflation. For simplicity, we therefore omit this trivial sixth contribution to $\mathcal{S}_k$ in this thesis.

interaction between the inflaton field and the gauge field, our expressions align with those given in Ref. [433].

8.1.2 Vacuum and induced curvature perturbations

Let us next discuss a general strategy to solving the mode equation for ζ in Eq. (8.17). In principle, solving it requires one to treat the source functions, F_ρ , F_q , F_σ etc., appearing in this equation on an equal footing to ζ , that is, solving Eq. (8.17) alongside the equations of motions for the F_α 's in Eq. (8.14) and analogous equations for $\delta E \cdot B$. This is similar to the case of adiabatic and isocurvature perturbations in multifield inflationary models, see for instance Refs. [155, 434, 435]. In this way, one could account for the backreaction of metric perturbations on gauge-field modes. In our work, we however assume that this effect is negligible with respect to the tachyonic amplification of the gauge-field mode functions via the inflaton field. Following this assumption, we have that the gauge-field source terms in Eq. (8.19) can be entirely expressed in terms of the gauge-mode functions which have been computed from the homogeneous background. This aligns with the conventional approach in other works on scalar perturbations from axion inflation (see, e.g., Refs. [56, 57, 339]).

Applying this strategy, we have that Eq. (8.17) is given by an inhomogeneous differential equation with a source term which is entirely determined from background dynamics. A general solution to Eq. (8.17) can therefore be written as a homogeneous and a particular solution [57]

$$\zeta_{\mathbf{k}}(\eta) = \zeta_{\mathbf{k}}^{\text{vac}}(\eta) + \zeta_{\mathbf{k}}^{\text{ind}}(\eta). \quad (8.21)$$

The first term, $\zeta_{\mathbf{k}}^{\text{vac}}$, corresponds to the homogeneous solution and physically reflects the contribution to $\zeta_{\mathbf{k}}$ coming from vacuum fluctuations of the inflaton field and the spacetime metric. This term is therefore analogous to the standard term that is also present in the case of slow-roll inflation. The second term, $\zeta_{\mathbf{k}}^{\text{ind}}(\eta)$, is the particular solution and reflects the gauge-field-induced component of $\zeta_{\mathbf{k}}$.

We first study the vacuum piece. We follow the standard approach for inflation and treat $\zeta_{\mathbf{k}}^{\text{vac}}$ as a scalar quantum field, which we decompose into a mode function, and a set of creation and annihilation operators, $\hat{b}_{\mathbf{k}}^\dagger$ and $\hat{b}_{\mathbf{k}}$,

$$\zeta_{\mathbf{k}}^{\text{vac}}(\eta) = u(\eta, k) \hat{b}_{\mathbf{k}} + u^*(\eta, k) \hat{b}_{-\mathbf{k}}^\dagger. \quad (8.22)$$

Here, the mode functions solve the homogeneous differential equation

$$\mathcal{D}_k u(\eta, k) \equiv u''(\eta, k) + p_k(\eta) u'(\eta, k) + q_k(\eta) u(\eta, k) = 0, \quad (8.23)$$

and the ladder operators $\hat{b}_{\mathbf{k}}^\dagger$ and $\hat{b}_{\mathbf{k}}$ obey canonical commutation relations

$$[\hat{b}_{\mathbf{k}}, \hat{b}_{\mathbf{k}'}^\dagger] = \delta^{(3)}(\mathbf{k} - \mathbf{k}'), \quad [\hat{b}_{\mathbf{k}}, \hat{b}_{\mathbf{k}'}] = [\hat{b}_{\mathbf{k}}^\dagger, \hat{b}_{\mathbf{k}'}^\dagger] = 0. \quad (8.24)$$

In general, the mode equation in Eq. (8.23) is a second-order differential equation with real coefficients p_k and q_k , thus we can find two linearly independent solutions to it, $u_1(\eta, k)$ and $u_2(\eta, k)$. The mode function u is then a superposition of these two solutions,

$$u(\eta, k) = C_k^{(1)} u_1(\eta, k) + C_k^{(2)} u_2(\eta, k). \quad (8.25)$$

Here, $C_k^{(1,2)}$ are two complex coefficients that are determined by defining appropriate initial conditions for $u(\eta, k)$. We shall return to the discussion of initial conditions in the following

sections when we study scalar perturbations in a more concrete physical situation.

Next, let us consider ζ_k^{ind} . This piece can be determined by means of the Green function method (see, e.g., Refs. [Green:2009ds, Barnaby:2009mc, 52, 55]). The Green function $G_k(\eta, \tau)$ is defined as the solution to the equation

$$\mathcal{D}_k G_k(\eta, \tau) = \delta(\eta - \tau). \quad (8.26)$$

We can construct the retarded Green function from the two linearly independent solutions $u_1(\eta, k)$ and $u_2(\eta, k)$ to the homogeneous differential equation in Eq. (8.23). It is given by

$$G_k(\eta, \tau) = \frac{u_2(\eta, k)u_1(\tau, k) - u_1(\eta, k)u_2(\tau, k)}{W_k(\tau)} \theta(\eta - \tau). \quad (8.27)$$

The denominator in this expression is the Wronskian of u_1 and u_2 ,

$$W_k(\tau) = u_1(\tau, k)u_2'(\tau, k) - u_2(\tau, k)u_1'(\tau, k). \quad (8.28)$$

One may readily verify that $G_k(\eta, \tau)$ in Eq. (8.27) indeed obeys the retarded boundary conditions

$$G_k(\eta, \tau \geq \eta) \equiv 0, \quad \left. \frac{\partial}{\partial \eta} G_k(\eta, \tau) \right|_{\tau \rightarrow \eta} = 1. \quad (8.29)$$

Finally, the induced contribution, ζ_k^{ind} , is then given by

$$\zeta_k^{\text{ind}}(\eta) = \int_{-\infty}^{\eta} d\tau G_k(\eta, \tau) \mathcal{S}_k(\tau), \quad (8.30)$$

where $\mathcal{S}_k(\eta)$ is the source term appearing on the right-hand side of the inhomogeneous differential equation Eq. (8.17) as given in Eq. (8.19).

One can now find a more useful expression the source term by expressing the F_α 's and $\delta_{\mathbf{E}, \mathbf{B}}$ in terms of the gauge-field operators $\mathbf{E}$ and $\mathbf{B}$. As outlined in the beginning of this section, we do this by evaluating the gauge-field operators on the homogeneous and isotropic FLRW background. In this case, we can recycle our discussion of the quantized gauge field during axion inflation in chapter 3; we write the gauge potential in temporal gauge, $A_\mu = (0, -\mathbf{A})$ with an additional Coulomb constraint $\text{div } \mathbf{A} = 0$, such that the electric- and magnetic-field three-vectors are related to the vector potential through $\mathbf{E} = -\mathbf{A}'/a^2$ and $\mathbf{B} = \text{rot } \mathbf{A}/a^2$ (see Eq. (3.10a)). The quantized gauge field can then be decomposed into its own set of creation and annihilation operators, $\hat{a}_{\mathbf{k}, \lambda}^\dagger$ and $\hat{a}_{\mathbf{k}, \lambda}$, and mode functions $A_\lambda(\eta, k)$ with different momenta $\mathbf{k}$ and circular polarizations λ as in Eq. (3.15). These mode functions obey the standard mode equation for axion inflation, namely

$$A_\lambda''(\eta, k) + [k^2 - 2\lambda k \xi(\eta) \mathcal{H}(\eta)] A_\lambda(\eta, k) = 0, \quad (8.31)$$

with the tachyonic instability parameter ξ defined in Eq. (3.20).

The quantized electric- and magnetic-field operators can then be expressed in Fourier space as

$$\mathbf{E}(\eta, \mathbf{x}) = \int \frac{d^3 \mathbf{k}}{(2\pi)^{3/2}} \sum_{\lambda=\pm} \left[\epsilon^\lambda(\mathbf{k}) \hat{a}_{\mathbf{k}, \lambda} \mathcal{E}_\lambda(\eta, k) e^{i\mathbf{k} \cdot \mathbf{x}} + \text{h.c.} \right], \quad (8.32a)$$

$$\mathbf{B}(\eta, \mathbf{x}) = \int \frac{d^3 \mathbf{k}}{(2\pi)^{3/2}} \sum_{\lambda=\pm} \left[\epsilon^\lambda(\mathbf{k}) \hat{a}_{\mathbf{k}, \lambda} \mathcal{B}_\lambda(\eta, k) e^{i\mathbf{k} \cdot \mathbf{x}} + \text{h.c.} \right], \quad (8.32b)$$

with the respective mode functions given by

$$\mathcal{E}_\lambda(\eta, k) = -\frac{1}{a^2(\eta)} A'_\lambda(\eta, k), \quad \mathcal{B}_\lambda(\eta, k) = \frac{\lambda k}{a^2(\eta)} A_\lambda(\eta, k). \quad (8.33)$$

As before, the polarization vectors $\epsilon^\lambda(\mathbf{k})$ obey Eq. (3.16), and $\hat{a}_{\mathbf{k},\lambda}^\dagger$ ($\hat{a}_{\mathbf{k},\lambda}$) are defined to be canonically commuting. Furthermore, we have that the ladder operators for the gauge field are commuting with those of $\zeta_{\mathbf{k}}^{\text{vac}}$ since the two are independent fields:

$$[\hat{b}_{\mathbf{k}}, \hat{a}_{\mathbf{k}',\lambda}] = [\hat{b}_{\mathbf{k}}, \hat{a}_{\mathbf{k}',\lambda}^\dagger] = 0. \quad (8.34)$$

The source term $\mathcal{S}_{\mathbf{k}}(\eta)$ is then quadratic in the gauge-field operators $\mathbf{E}$ and $\mathbf{B}$, and can be written in the form

$$\begin{aligned} \mathcal{S}_{\mathbf{k}}(\eta) = \int \frac{d^3\mathbf{p}}{(2\pi)^{3/2}} \sum_{\lambda,\lambda'=\pm} & \left[K_{\lambda,\mathbf{p};\lambda',\mathbf{k}-\mathbf{p};\mathbf{k}}(\eta) \hat{a}_{\mathbf{p},\lambda} \hat{a}_{\mathbf{k}-\mathbf{p},\lambda'} + K_{\lambda,\mathbf{p};\lambda',\mathbf{k}+\mathbf{p};\mathbf{k}}^*(\eta) \hat{a}_{\mathbf{p},\lambda}^\dagger \hat{a}_{-\mathbf{k}-\mathbf{p},\lambda'}^\dagger \right. \\ & \left. + L_{\lambda',\mathbf{p}-\mathbf{k};\lambda,\mathbf{p};\mathbf{k}}(\eta) \hat{a}_{\mathbf{p},\lambda} \hat{a}_{\mathbf{p}-\mathbf{k},\lambda'}^\dagger + L_{\lambda,\mathbf{p};\lambda',\mathbf{k}+\mathbf{p};\mathbf{k}}(\eta) \hat{a}_{\mathbf{p},\lambda}^\dagger \hat{a}_{\mathbf{k}+\mathbf{p},\lambda'} \right], \end{aligned} \quad (8.35)$$

The functions K and L are scalar functions of both the polarization vectors and the mode functions $\mathcal{E}_\lambda(\eta, k)$ and $\mathcal{B}_\lambda(\eta, k)$ and are linearly dependent on the coefficients $r^{(\alpha)}$ in Eq. (8.17). Once again, we defer the lengthy expressions for K and L to App. D. The fact that only two functions appear in front of the quadratic combinations of creation and annihilation operators in Eq. (8.35) follows on general grounds, since creation and annihilation operators commute if $\mathbf{k} \neq 0$ and the source $\mathcal{S}_{\mathbf{k}}(\eta)$ is a Hermitian operator.

8.1.3 The scalar power spectrum and bispectrum

When studying the production of stochastic quantum fields such as $\zeta_{\mathbf{k}}$, it is desirable to express our final results in terms of statistically meaningful quantities. As introduced in Sec. 1.4.1, the predictions for inflationary perturbations are conveniently expressed in terms of power spectra, which relate to the amplitude of their two-point correlation function. The amplitude of the ζ power spectrum is then constrained by CMB measurements, as described in Sec. 1.5. In this chapter, we are also concerned with studying deviations from Gaussian perturbations. These deviations are, e.g., signaled by a non-vanishing three-point correlation function of $\zeta_{\mathbf{k}}$. Analogous to the power spectrum, its amplitude is described by the so-called bispectrum, which is then constrained by CMB observations [15, 49].

The scalar power spectrum

We first study the scalar power spectrum, which we define according to our prescription in chapter 1,

$$\langle \zeta_{\mathbf{k}}(\eta) \zeta_{\mathbf{k}'}(\eta) \rangle = \frac{2\pi^2}{k^3} \mathcal{P}_\zeta(k, \eta) \delta(\mathbf{k} + \mathbf{k}'). \quad (8.36)$$

The two-point correlation function on the left-hand side can then be evaluated based on the general solution of $\zeta_{\mathbf{k}}$ described in the previous section. This means, $\zeta_{\mathbf{k}}$ splits into two terms, one corresponding to the vacuum contribution, and one which is sourced by gauge fields. The two-point correlation function for $\zeta_{\mathbf{k}}$ also splits into two, since the vacuum and induced components

are statistically independent;⁴

$$\langle \zeta_{\mathbf{k}}(\eta) \zeta_{\mathbf{k}'}(\eta) \rangle = \langle \zeta_{\mathbf{k}}^{\text{vac}}(\eta) \zeta_{\mathbf{k}'}^{\text{vac}}(\eta) \rangle + \langle \zeta_{\mathbf{k}}^{\text{ind}}(\eta) \zeta_{\mathbf{k}'}^{\text{ind}}(\eta) \rangle. \quad (8.37)$$

Therefore, also the scalar power spectrum separates into a vacuum and an induced contribution,

$$\mathcal{P}_{\zeta}(\eta, k) = \mathcal{P}_{\zeta}^{\text{vac}}(\eta, k) + \mathcal{P}_{\zeta}^{\text{ind}}(\eta, k). \quad (8.38)$$

The two terms can be directly computed starting from the expressions for $\zeta_{\mathbf{k}}^{\text{vac}}$ and $\zeta_{\mathbf{k}}^{\text{ind}}$ in Eqs. (8.22) and (8.30) and the explicit expression for the source function in Eq. (8.35). Using Wick's theorem to evaluate the expectation values of creation and annihilation operators in Eq. (8.22) and Eq. (8.35), one obtains

$$\mathcal{P}_{\zeta}^{\text{vac}}(\eta, k) = \frac{k^3}{2\pi^2} |u(\eta, k)|^2, \quad (8.39a)$$

$$\mathcal{P}_{\zeta}^{\text{ind}}(\eta, k) = \frac{k^3}{\pi^2} \int \frac{d^3 \mathbf{p}}{(2\pi)^3} \sum_{\lambda, \lambda' = \pm} \left| \int_{-\infty}^{\eta} d\tau G_k(\eta, \tau) K_{\lambda, \mathbf{p}; \lambda', \mathbf{k}-\mathbf{p}; \mathbf{k}}(\tau) \right|^2. \quad (8.39b)$$

In the expression for $\mathcal{P}_{\zeta}^{\text{ind}}$, only the terms proportional to the function K in $\mathcal{S}_{\mathbf{k}}$ survives, as the terms proportional to L either contain at least one annihilation operator $\hat{a}_{\lambda, \mathbf{k}}$ acting on the Bunch–Davies vacuum state $|0\rangle$ from the left, or a creation operator $\hat{a}_{\lambda, \mathbf{k}}^{\dagger}$ acting on $\langle 0|$ from the right.

The bispectrum

Similar to the scalar power spectrum, we can also derive a general expression for the scalar bispectrum by applying the expressions for $\zeta_{\mathbf{k}}^{\text{vac}}$ and $\zeta_{\mathbf{k}}^{\text{ind}}$ in Eqs. (8.22) and (8.30). We define the bispectrum as

$$\langle \zeta_{\mathbf{k}_1}(\eta) \zeta_{\mathbf{k}_2}(\eta) \zeta_{\mathbf{k}_3}(\eta) \rangle = \delta(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) \mathcal{B}_{\zeta}(k_1, k_2, k_3). \quad (8.40)$$

Note that this quantity also depends on time, η ; however, for brevity, we do not write this dependence explicitly in the arguments.

At linear order in perturbation theory, the only non-zero three-point correlator comes from the induced contribution,

$$\langle \zeta_{\mathbf{k}_1}(\eta) \zeta_{\mathbf{k}_2}(\eta) \zeta_{\mathbf{k}_3}(\eta) \rangle = \langle \zeta_{\mathbf{k}_1}^{\text{ind}}(\eta) \zeta_{\mathbf{k}_2}^{\text{ind}}(\eta) \zeta_{\mathbf{k}_3}^{\text{ind}}(\eta) \rangle. \quad (8.41)$$

The three-point correlator for the vacuum contribution vanishes because of Wick's theorem, as it necessarily contains three creation and/or annihilation operators, one of which annihilates either $|0\rangle$ or $\langle 0|$.⁵

⁴Note that the vacuum and induced contributions are also statistically independent for classical fields, see, for instance, Sec 5.6 in Ref. [436].

⁵This is only true at linear order in perturbation theory. At higher order in perturbation theory, inflaton self-interaction also leads to a non-zero vacuum bispectrum. The amplitude of this bispectrum is however slow-roll suppressed [211, 212, 437].

Using Eq. (8.35), the bispectrum is then explicitly given by

$$\begin{aligned} \mathcal{B}_\zeta(k_1, k_2, k_3) = & \left(\frac{2}{\pi}\right)^{3/2} \int \frac{d^3\mathbf{p}}{(2\pi)^3} \sum_{\lambda_1, \lambda_2, \lambda_3 = \pm 1} \int_{-\infty}^{\eta} d\tau_1 G_{k_1}(\eta, \tau_1) K_{\lambda_1, \mathbf{p}; \lambda_2, \mathbf{k}_1 - \mathbf{p}; \mathbf{k}_1}(\tau_1) \\ & \times \int_{-\infty}^{\eta} d\tau_2 G_{k_2}(\eta, \tau_2) L_{\lambda_2, \mathbf{k}_1 - \mathbf{p}; \lambda_3, -\mathbf{k}_3 - \mathbf{p}; \mathbf{k}_2}(\tau_2) \int_{-\infty}^{\eta} d\tau_3 G_{k_3}(\eta, \tau_3) K_{\lambda_3, -\mathbf{k}_3 - \mathbf{p}; \lambda_1, \mathbf{p}; -\mathbf{k}_3}^*(\tau_3). \end{aligned} \quad (8.42)$$

Statistical homogeneity implies that the bispectrum is only non-vanishing when the external momenta $\mathbf{k}_1$, $\mathbf{k}_2$, and $\mathbf{k}_3$ form a triangle configuration with $\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 = 0$. Because of statistical isotropy, the orientation of this triangle in space is irrelevant, and the bispectrum depends only on the shape of the triangle. This shape can be conveniently parametrized by the moduli of its sides, k_1 , k_2 , and k_3 . These satisfy the triangle inequality

$$k_l + k_m \geq k_n, \quad (8.43)$$

with (lmn) running over all permutations of the indices 1, 2, and 3. Although the bispectrum in Eq. (8.42) depends on the three-vectors $\mathbf{k}_1$, $\mathbf{k}_2$, and $\mathbf{k}_3$, the dependence on their orientation will vanish once the integral over the internal momentum $\mathbf{p}$ in Eq. (8.42) is performed. Furthermore, the bispectrum is symmetric under a permutation of the external momenta, as one can always perform a spatial rotation to bring the resulting triangle back into its original form. Therefore, we can consider the fixed configuration $k_1 \geq k_2 \geq k_3$ without loss of generality.

The bispectrum evidently has a more complicated structure than the scalar power spectrum. Just as for the power spectrum, one can define an overall amplitude for the bispectrum evaluated at a characteristic momentum scale k . For instance, one can consider the bispectrum's amplitude in the equilateral configuration, $k_1 = k_2 = k_3 = k$,

$$\mathcal{B}_\zeta^{\text{eq}}(k) \equiv \mathcal{B}_\zeta(k, k, k). \quad (8.44)$$

Furthermore, for a fixed momentum scale $k \equiv \max\{k_1, k_2, k_3\} = k_1$, one can describe the dependence of the bispectrum on the triangle configuration of the external momenta by means of a dimensionless shape function

$$\Sigma(k; x_2, x_3) \equiv \mathcal{N}^2(k_1 k_2 k_3) \mathcal{B}_\zeta(k_1, k_2, k_3) \quad (8.45)$$

where $x_2 = k_2/k$, $x_3 = k_3/k$ and $\mathcal{N}$ is a normalization factor which is subject to convention. For instance, it can be chosen such that the shape function is normalized to unity in the equilateral configuration, $\Sigma(k; 1, 1) = 1$, which is the convention we shall employ. It should be noted that there are different conventions for the overall momentum scale k in the literature, so the choice of arguments for the shape function is not unique, see for instance Refs. [438, 439]. The convention we rely on is the one given in Ref. [438].

CMB constraints on primordial scalar non-Gaussianity are only known for a few particular triangle configurations of the bispectrum, for instance the local [437, 440], equilateral [441], or orthogonal [442] shapes. In our convention, where the normalization factor $\mathcal{N}$ in Eq. (8.45) is

defined with respect to the equilateral shape, these three shape functions are

$$\Sigma_{\text{local}}(x_2, x_3) = \frac{1 + x_2^3 + x_3^3}{3x_2x_3}, \quad (8.46a)$$

$$\Sigma_{\text{equil}}(x_2, x_3) = \frac{(1 + x_2 - x_3)(1 - x_2 + x_3)(x_2 + x_3 - 1)}{x_2x_3}, \quad (8.46b)$$

$$\Sigma_{\text{ortho}}(x_2, x_3) = 3\Sigma_{\text{equil}}(x_2, x_3) - 2. \quad (8.46c)$$

To compare the predictions for the bispectrum in a given model to CMB constraints, the bispectrum can be parametrized in terms of a unique, model-dependent shape function Σ (not necessarily one of the three patterns given above). To quantify the similarity between this model-dependent shape function and the standard templates in Eq. (8.46) it is then convenient to define a scalar product on the space of shape functions, which can be used to compute the cosine of the “angle” between two shape functions Σ and Σ' . The definition of this scalar product is again subject to different conventions, see Refs. [438, 439]. Once again, for sake of simplicity, we rely on the convention given in Ref. [438], and define the scalar product between two shapes Σ and Σ' as

$$(\Sigma, \Sigma') = \int dx_2 dx_3 \Sigma(x_2, x_3) \Sigma'(x_2, x_3). \quad (8.47)$$

Here, the integration is performed over the region constrained by the inequalities

$$x_2 + x_3 \geq 1, \quad x_3 \leq x_2 \leq 1, \quad (8.48)$$

which follow from the triangle inequality and our ordering of the variables, $k_1 \geq k_2 \geq k_3$. The cosine of the generalized angle $\Theta_{\Sigma, \Sigma'}$ between two shapes in the vector space of shape functions is then given by

$$\cos \Theta_{\Sigma, \Sigma'} = \frac{(\Sigma, \Sigma')}{\sqrt{(\Sigma, \Sigma)(\Sigma', \Sigma')}}. \quad (8.49)$$

With the general expressions for the power spectrum in Eq. (8.39), and the bispectrum, Eq. (8.42), including our convention for the bispectrum amplitude and shape, we have gathered all necessary ingredients which we need to apply our formalism to a concrete physics application. A first such application is the focus of the next section.

8.2 First application: Axion inflation without backreaction

For the first concrete application of our new formalism, we consider the case of axion inflation without backreaction effects, i.e., we assume that gauge fields are produced, but they do not disrupt the slow-roll evolution of the inflaton field. This assumption allows for a (semi)analytical computation of the scalar power spectrum and bispectrum, as has been discussed in detail in, e.g., Ref. [57], for the case where the gauge field only sources inflaton perturbations, and not metric perturbations. This calculation, for instance, yields the established analytical formula for the scalar power spectrum that we have given in Eq. (3.33). In this section, we follow this exact derivation to arrive at the functionally same expression for the scalar power spectrum, and a similar one for the scalar bispectrum. However, we compute these results using our new source function $\mathcal{S}_k$ in Eq. (8.19), which accounts for both sourced metric and inflaton perturbations. Our calculation therefore will yield a different amplitude for the scalar power spectrum and bispectrum, which we can then compare to the original results in Ref. [57] to assess the importance of gauge-field-induced metric perturbations.

To apply our formalism in this analytical case study, we need to make an important simplification, which also appropriately reflects the assumption of weak gauge-field backreaction. Below Eq. (8.13), we have remarked upon the fact that metric perturbations are sourced directly from the gauge-field source functions F_α defined in Eq. (8.12), but may also couple to backreaction terms such as $\langle \mathbf{E}^2 \rangle$, $\langle \mathbf{B} \rangle^2$, $\langle (\mathbf{E} \cdot \mathbf{B}) \rangle$ etc. The assumption of weak backreaction entails that such terms are small compared to, e.g., the inflaton's potential energy. For the sake of our analytical calculation, this implies setting these terms strictly to zero when computing the scalar power spectrum. In other words, the comparison we perform here only accounts for the sourcing of scalar metric perturbations via the source functions F_α , while neglecting the backreaction of gauge fields on ζ_k .

8.2.1 Taking inventory

We begin our derivation by explicitly laying out the assumption for this section. Then, we go through the individual ingredients that enter into the general expressions for the scalar power spectrum and bispectrum in Eqs. (8.39) and (8.42), i.e., the scalar mode functions, $u(\eta, k)$, the Green function, $G_k(\eta, \tau)$, and the source function $\mathcal{S}_k$, to establish explicit expressions for these objects for the simple dynamical setup we consider here.

As explained above, the calculation we perform in this section assumes absent or negligible gauge-field backreaction. For our formalism, this entails setting

$$\langle \mathbf{E}^2 \rangle \approx 0, \quad \langle \mathbf{B}^2 \rangle \approx 0, \quad \langle (\mathbf{E} \cdot \mathbf{B}) \rangle \approx 0, \quad (8.50)$$

in all equations for the coefficients $r^{(\alpha)}$, p_k , and q_k appearing in Eq. (8.17). The left-hand side (LHS) of Eq. (8.17) then takes the form of the Mukhanov–Sasaki equation, as we have given it in Eq. (1.75),

$$\text{LHS} = \zeta_k'' + 2\frac{z'}{z}\zeta_k' + k^2\zeta_k, \quad (8.51)$$

with $z = a\varphi'/\mathcal{H}$. The expressions for the coefficients $r^{(\alpha)}$ appearing on the right-hand side of Eq. (8.17) for this case are given in App. D.

In the case of negligible gauge-field backreaction, we may assume that the inflaton field is on the slow-roll trajectory (cf. Sec. 3.3), such that $\dot{\varphi}$ is nearly constant, as is the Hubble rate H . Corrections to these are of the order of the slow-roll parameters ε_V and η_V (see Eq. (1.30)), which we neglect in our computation. This means, we work under the idealized assumption of perfect de Sitter space, implying the exact relation

$$a(\eta) = -\frac{1}{H\eta}. \quad (8.52)$$

Furthermore, we have that the instability parameter ξ for axion inflation is also constant

$$\xi = \frac{\beta}{M_P} \frac{\dot{\varphi}}{2H} = \text{const}. \quad (8.53)$$

The results of our analytical calculation are then parametrized in terms of H , ξ and the inflaton velocity, $\dot{\varphi}$.

Scalar mode functions and the Green function

To compute the power spectrum and the bispectrum, we need to determine both the homogeneous and inhomogeneous parts of ζ_k as outlined in Sec. 8.1.

At first, we consider the homogeneous (vacuum) part, $\zeta_{\mathbf{k}}^{\text{vac}}$, which we decompose into a mode function and ladder operators as in Eq. (8.23). In the case of absent gauge-field backreaction, the mode functions $u(\eta, k)$ solves

$$u''(\eta, k) + 2\frac{z'(\eta)}{z(\eta)}u'(\eta, k) + k^2u(\eta, k) = 0. \quad (8.54)$$

This is the standard Mukhanov–Sasaki equation one also needs to solve to determine the scalar power spectrum in the case of single-field slow-roll inflation. To solve it, we first remove the term $2(z'/z)u'$ by introducing the rescaled mode function $v(\eta, k) = z(\eta)u(\eta, k)$. The modes $v(\eta, k)$ then solve

$$v''(\eta, k) + \left(k^2 - \frac{z''(\eta)}{z(\eta)}\right)v(\eta, k) = 0. \quad (8.55)$$

This can be further simplified by using that $z \propto a \propto \eta^{-1}$ for constant H and constant $\dot{\phi}$, yielding

$$v''(\eta, k) + \left(k^2 - \frac{2}{\eta^2}\right)v(\eta, k) = 0. \quad (8.56)$$

This is precisely the mode equation for a massless scalar field in de Sitter space, as we have discussed it in chapter 1. In the limit $-k\eta \gg 1$, we can therefore select Bunch–Davies boundary conditions for $v(\eta, k)$, which then uniquely determines $u(\eta, k)$. The exact solution for $u(\eta, k)$ which obeys these boundary conditions can be expressed in terms of Hankel functions and follows from Eq. (1.62). It is given by

$$u(\eta, k) = \frac{H^2}{\dot{\phi}} \frac{\sqrt{\pi}}{2} (-\eta)^{3/2} H_{3/2}^{(1)}(-k\eta) = -\frac{H^2}{\dot{\phi}} \frac{1}{\sqrt{2}k^{3/2}} (-k\eta + i)e^{-ik\eta}. \quad (8.57)$$

Since the coefficients of the differential equation in Eq. (8.54) are real, we furthermore have that $u(\eta, k)$ and its complex conjugate constitute two linearly independent solutions to it. We can then use these to also determine an analytical expression for the retarded Green function by means of Eq. (8.27). The result one obtains is

$$G_{\mathbf{k}}(\eta, \eta') = \theta(y' - y) \frac{1}{ky'^2} [(y - y') \cos(y - y') - (1 + yy') \sin(y - y')], \quad (8.58)$$

with $y = -k\eta$ and $y' = -k\eta'$. We are interested in computing the scalar power spectrum and bispectrum for modes which have left the horizon at time η , i.e., $y = -k\eta \ll 1$. This allows one to simplify the Green function in Eq. (8.58) to

$$G_{\mathbf{k}}(\eta, \eta') \simeq \theta(y' - y) \frac{1}{ky'^2} (\sin y' - y' \cos y'). \quad (8.59)$$

It is this second, approximate form of the retarded Green function which we shall use in what follows.

Gauge-field mode functions

To compute the induced power spectrum and bispectrum, we also need the gauge-field mode functions $A_\lambda(\eta, k)$ entering into the source function $\mathcal{S}_{\mathbf{k}}$ in Eq. (8.35). In the case of constant ξ and

H , the mode equation in Eq. (8.31) becomes

$$A_\lambda(\eta, k)'' + \left(k^2 + \lambda k \frac{2\xi}{\eta}\right) A_\lambda(\eta, k) = 0. \quad (8.60)$$

As we have discussed in chapter 3, this equation can be solved exactly in terms of Whittaker functions, $W_{\kappa, \mu}$ as given in Eq. (3.28). This implies that the mode functions for the electric and magnetic field in Eq. (8.33) are given by

$$\mathcal{E}_\lambda(\eta, k) = -\frac{1}{a^2} \sqrt{\frac{k}{2}} \mathcal{U}_\lambda(-k\eta), \quad \mathcal{B}_\lambda(\eta, k) = \frac{\lambda}{a^2} \sqrt{\frac{k}{2}} \mathcal{W}_\lambda(-k\eta), \quad (8.61)$$

with

$$\mathcal{W}_\lambda(\rho) = e^{\lambda\pi\xi/2} W_{-i\lambda\xi, 1/2}(-2i\rho) \quad (8.62)$$

and

$$\mathcal{U}_\lambda(\rho) = e^{\lambda\pi\xi/2} \frac{1}{\rho} \left[i(\rho - \lambda\xi) W_{-i\lambda\xi, 1/2}(-2i\rho) + W_{1-i\lambda\xi, 1/2}(-2i\rho) \right], \quad (8.63)$$

where the second relationship follows from the recursion relationship for derivatives of Whittaker functions, see App. A.

The source term

The source term $\mathcal{S}_k(\eta)$ in Eq. (8.35) is entirely given by the functions K and L , which we also need to compute the power spectrum and bispectrum in Eqs. (8.39b) and (8.42). For the simplifying assumptions made in this section, these functions are given by

$$\begin{aligned} K_{\lambda, \mathbf{p}; \lambda', \mathbf{p}'; \mathbf{k}}(\eta) = & (\epsilon_\lambda(\mathbf{p}) \cdot \epsilon_{\lambda'}(\mathbf{p}')) \frac{a^2 \mathcal{H}^2}{\varphi'^2} \left\{ \xi [1 + h_1] \left[\mathcal{E}_\lambda(\eta, p) \mathcal{B}_{\lambda'}(\eta, p') + \mathcal{B}_{\lambda, \mathbf{p}}(\eta) \mathcal{E}_{\lambda', \mathbf{p}'}(\eta) \right] \right. \\ & + h_2 \left[\mathcal{E}_\lambda(\eta, p) \mathcal{E}_{\lambda'}(\eta, p') + \mathcal{B}_\lambda(\eta, p) \mathcal{B}_{\lambda'}(\eta, p') \right] \\ & \left. + h_3 \frac{k}{\mathcal{H}} \left[\mathcal{E}_\lambda(\eta, p) \mathcal{B}_{\lambda'}(\eta, p') - \mathcal{B}_\lambda(\eta, p) \mathcal{E}_{\lambda'}(\eta, p') \right] \right\}, \quad (8.64a) \end{aligned}$$

$$\begin{aligned} L_{\lambda, \mathbf{p}; \lambda', -\mathbf{p}'; \mathbf{k}}(\eta) = & (\epsilon_\lambda(\mathbf{p}) \cdot \epsilon_{\lambda'}(\mathbf{p}')) \frac{a^2 \mathcal{H}^2}{\varphi'^2} \left\{ \xi [1 + h_1] \left[\mathcal{E}_\lambda^*(\eta, p) \mathcal{B}_{\lambda'}(\eta, p') + \mathcal{B}_\lambda^*(\eta, p) \mathcal{E}_{\lambda'}(\eta, p') \right] \right. \\ & + h_2 \left[\mathcal{E}_\lambda^*(\eta, p) \mathcal{E}_{\lambda'}(\eta, p') + \mathcal{B}_\lambda^*(\eta, p) \mathcal{B}_{\lambda'}(\eta, p') \right] \\ & \left. + h_3 \frac{k}{\mathcal{H}} \left[\mathcal{E}_\lambda^*(\eta, p) \mathcal{B}_{\lambda'}(\eta, p') - \mathcal{B}_\lambda^*(\eta, p) \mathcal{E}_{\lambda'}(\eta, p') \right] \right\}, \quad (8.64b) \end{aligned}$$

where $\mathbf{p} + \mathbf{p}' = \mathbf{k}$, and $p = |\mathbf{p}|$, $p' = |\mathbf{p}'|$. The dimensionless factors h_i are given by

$$h_1 = -x^2, \quad h_2 \simeq 2x^2 + \mathcal{O}(\epsilon_V, \eta_V), \quad h_3 = -\frac{x}{2}(1 + x^2) + \mathcal{O}(\epsilon_V, \eta_V), \quad (8.65)$$

with

$$x \equiv \frac{(\lambda p - \lambda' p')}{k}. \quad (8.66)$$

Higher-order corrections to h_2 and h_3 in the slow-roll parameters ϵ_V and η_V can be neglected in our case, as we consider H and $\dot{\varphi}$ to be constant. Writing the factors K and L in terms of the factors h_i is convenient to compare our new results to the previous approach in the literature, where

metric perturbations have been neglected (for instance Ref. [57]). These old results are retrieved by setting $h_1 = h_2 = h_3 = 0$ in Eq. (8.64). In what follows, we denote quantities that have been derived in this case with a tilde, i.e., the source functions with the h_i 's set to zero are denoted by $\tilde{K}$ and $\tilde{L}$ etc.

We then introduce the following useful rescaled dimensionless momenta

$$\mathbf{p}_* \equiv \frac{1}{q} \mathbf{p}, \quad \mathbf{p}'_* \equiv \frac{1}{q} \mathbf{p}', \quad x \equiv \frac{|\mathbf{k}|}{q}, \quad (8.67)$$

Here, q is an overall momentum scale we keep arbitrary for the moment. Using these new variables alongside the expressions for the Green function in Eq. (8.59) and the ones for L and K in Eq. (8.64), one can then rewrite the conformal time integrals appearing in both the induced power spectrum in Eq. (8.39b) and the bispectrum in Eq. (8.42). The first such integral can be written as

$$\int_{-\infty}^{\eta} d\tau G_k(\eta, \tau) K_{\lambda, \mathbf{p}; \lambda', \mathbf{p}'; \mathbf{k}}(\tau) = -\frac{H^4}{2q^3 \dot{\varphi}^2} (\epsilon_{\lambda}(\mathbf{p}_*) \cdot \epsilon_{\lambda'}(\mathbf{p}'_*)) \frac{\sqrt{p_* p'_*}}{x^4} \mathcal{I}(\lambda, p_*; \lambda', p'_*; x). \quad (8.68)$$

Rewriting the integral over τ in terms of $y = -k\tau$, the dimensionless integral $\mathcal{I}$ is

$$\begin{aligned} \mathcal{I}(\lambda, p_*; \lambda', p'_*; x) = & \int_0^{y_{\text{UV}}} dy (\sin y - y \cos y) \left\{ (\xi + \xi h_1 + y h_3) \lambda' \mathcal{U}_{\lambda} \left(\frac{p_*}{x} y \right) \mathcal{W}_{\lambda'} \left(\frac{p'_*}{x} y \right) \right. \\ & + (\xi + \xi h_1 - y h_3) \lambda \mathcal{W}_{\lambda} \left(\frac{p_*}{x} y \right) \mathcal{U}_{\lambda'} \left(\frac{p'_*}{x} y \right) \\ & \left. - h_2 \left[\mathcal{U}_{\lambda} \left(\frac{p_*}{x} y \right) \mathcal{U}_{\lambda'} \left(\frac{p'_*}{x} y \right) + \lambda \lambda' \mathcal{W}_{\lambda} \left(\frac{p_*}{x} y \right) \mathcal{W}_{\lambda'} \left(\frac{p'_*}{x} y \right) \right] \right\}. \quad (8.69) \end{aligned}$$

The integral over the source function L has exactly the same form as the one over the source function K with the only difference being in the functions $\mathcal{W}_{\lambda}(p_* y/x)$ and $\mathcal{U}_{\lambda}(p_* y/x)$, which have to be replaced by their complex conjugate. A simple way of denoting this is by adding a complex conjugation to the corresponding pair of momenta and polarizations appearing as arguments in the integral $\mathcal{I}$,

$$\int_{-\infty}^{\eta} d\tau G_k(\eta, \tau) L_{\lambda, -\mathbf{p}; \lambda', \mathbf{p}'; \mathbf{k}}(\tau) = -\frac{H^4}{2q^3 \dot{\varphi}^2} (\epsilon_{\lambda}(\mathbf{p}_*) \cdot \epsilon_{\lambda'}(\mathbf{p}'_*)) \frac{\sqrt{p_* p'_*}}{x^4} \mathcal{I}((\lambda, p_*)^*; \lambda', p'_*; x). \quad (8.70)$$

In principle, the integration over y in Eq. (8.69) should run over $-k\eta$ to infinity. However, our interest is in computing the power spectrum and bispectrum for modes that are super-horizon, $-k\eta \ll 1$. In practice, we can therefore set the lower integration boundary in $\mathcal{I}$ to zero. Furthermore, $\mathcal{I}$ is also UV-divergent, as we are integrating over gauge-field modes $A_{\lambda}(\eta, k)$ with arbitrarily large momenta. To treat this UV divergence, we follow the same approach as in previous chapters, which is to introduce a UV cutoff. As before, a suitable UV regulator is the instability scale k_{h} given in Eq. (3.24). In the particular case we are studying here, with constant and positive ξ , we have $k_{\text{h}} = 2\xi aH$. This is the same regularization procedure as was also used in previous computations of the induced power spectrum and bispectrum, see e.g., Ref. [55].⁶ However, since this choice of a cut-off scale is purely heuristic, we follow the example of Ref. [103], to also study

⁶A proper QFT regularization of the gauge-field-induced scalar power spectrum and bispectrum has been considered in Ref. [443] for the case of a simple kinetic-coupling model for de Sitter space. However, we continue to rely on the cutoff-regularization method, as this regularization procedure can also be readily applied to the case where the dynamics

how our results are affected by $\mathcal{O}(1)$ changes to the cutoff scale. We therefore set the following upper integration limit for the integral $\mathcal{I}$:

$$y_{\text{UV}} = \gamma_{\text{UV}} \min \left\{ \frac{x}{p_*}, \frac{x}{p'_*} \right\}, \quad (8.71)$$

where the additional parameter γ_{UV} accounts for the possible theoretical uncertainties in the choice of the cutoff. The reference value is $\gamma_{\text{UV}} = 2\xi$, but, in what follows, we shall also study the cases $\gamma_{\text{UV}} = 2\xi/3$ and $\gamma_{\text{UV}} = 6\xi$ corresponding approximately to a variation of the cutoff scale by an order of magnitude.

8.2.2 Results

Putting all ingredients from the previous section together, we can now compute both the scalar power spectrum and bispectrum.

The scalar power spectrum

Let us begin with the scalar power spectrum. Firstly, the vacuum contribution to $\mathcal{P}_\zeta(k)$ can be readily computed by substituting the mode function $u(\eta, k)$ in Eq. (8.57) into Eq. (8.39a). For super-horizon modes, $-k\eta \ll 1$, one retains the standard slow-roll result (see Eq. (1.77));

$$\mathcal{P}_\zeta^{\text{vac}}(\eta, k) = \left(\frac{H^2}{2\pi\dot{\phi}} \right)^2 (1 + (k\eta)^2)^{-k\eta \ll 1} \simeq \left(\frac{H^2}{2\pi\dot{\phi}} \right)^2. \quad (8.72)$$

Next, we turn towards the induced contribution. Following Refs. [56, 57], it may be parametrized as

$$\mathcal{P}_\zeta^{\text{ind}}(\eta, k) = e^{4\pi\xi} \left(\frac{H^2}{2\pi\dot{\phi}} \right)^4 f_2(\xi). \quad (8.73)$$

Using Eq. (8.68) with $k = |\mathbf{k}|$ as the momentum scale q , the functions $f_2(\xi)$ is given by

$$f_2(\xi) = \frac{e^{-4\pi\xi}}{8\pi} \int d^3\mathbf{p}_* \sum_{\lambda, \lambda' = \pm} \left(1 - \lambda\lambda' \frac{\mathbf{p}_* \cdot \mathbf{p}'_*}{p_* p'_*} \right)^2 p_* p'_* |\mathcal{I}(\lambda, p_*; \lambda', p'_*; 1)|^2, \quad (8.74)$$

with $\mathbf{p}'_* = \mathbf{k}/k - \mathbf{p}_*$ and $\mathcal{I}$ given in Eq. (8.69). We also used the following identity for the scalar product of the two polarization vectors in Eq. (8.68);

$$\epsilon_\lambda(\mathbf{p}) \cdot \epsilon_{\lambda'}(\mathbf{p}') = \frac{1}{2} \left(1 - \lambda\lambda' \frac{\mathbf{p} \cdot \mathbf{p}'}{pp'} \right). \quad (8.75)$$

This identity can be readily obtained by choosing a fixed reference frame for the polarization vectors and decomposing the momenta $\mathbf{p}$ and $\mathbf{p}'$ in radial coordinates. Details on this derivation can be found in our original article, Ref. [2], App. E.

Note that the induced power spectrum in Eq. (8.73) is independent of the momentum k . This is a consequence of the assumption of constant H and ξ , which implies that any scalar-perturbation mode is subjected to the same gauge-field background at every moment in time, and thus exits the horizon in the same state as any other mode. This same result also has been found in the seminal work on this subject, Ref. [57]. It should be noted that this is no longer true in a realistic inflationary

of axion inflation are treated as fully time dependent, as we shall do in the following section. In this case, more advanced regularization techniques are very challenging to realize.

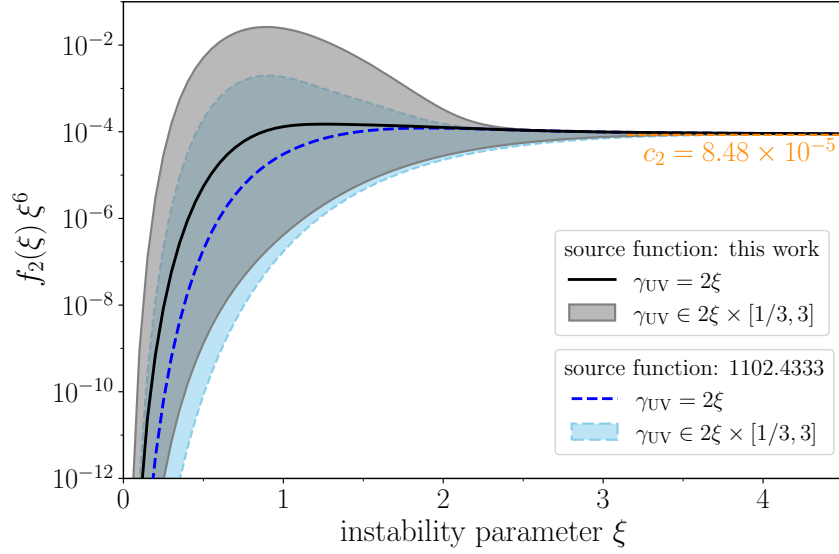


FIGURE 8.1: The function $f_2(\xi)$ parametrizing the gauge-field-induced part of the scalar power spectrum, multiplied by a factor ξ^6 . The black solid lines and gray-shaded regions represent the results of the full computation, which takes into account the metric perturbations, while the blue dashed line and light-blue-shaded region correspond to the function $\tilde{f}_2(\xi)$ in the case when the metric perturbations are neglected. The thick black and blue lines correspond to the UV cutoff in the integral (8.69) at the instability scale, $\gamma_{UV} = 2\xi$, while the shaded bands of the corresponding color show the variation of the final result under an increase or a decrease of the UV cutoff by a factor of 3. The orange dashed line starting around $\xi = 3$ shows the leading term in the expansion of $f_2(\xi)$ in the large- ξ approximation, see Eq. (8.76).

setting where the background is time-dependent. There, the power spectrum is k -dependent as suggested by the generic expression in Eq. (8.39). We will see this explicitly in the next section.

Continuing with our present analysis, one can obtain a simple analytical result for $f_2(\xi)$ by using an approximate form of the Whittaker function appearing in Eqs. (8.62) and (8.63). Neglecting the exponentially suppressed negative-helicity mode, and approximating the positive-helicity mode using Eq. (3.30), one can determine $f_2(\xi)$ as a series of inverse powers in ξ which approximately holds for $\xi \gg 1$. The leading-order term in this series behaves as (see also, e.g., Ref. [57])

$$f_2(\xi) \simeq \frac{c_2}{\xi^6}. \quad (8.76)$$

Using our formalism, we obtain the coefficient $c_2 \approx 8.48 \times 10^{-5}$. Using instead the source term from Ref. [57], which neglects gauge-field-sourced metric perturbations, one gets $\tilde{c}_2 \approx 7.47 \times 10^{-5}$, which is about 12 % smaller than our result. For details on this derivation, see App. F in our original work, Ref. [2].

We also compute the function $f_2(\xi)$ numerically for different values of ξ , allowing us to go beyond the large- ξ approximation.⁷ We show the result of this computation in Fig. 8.1, wherein we plot $f_2(\xi)\xi^6$ as a function of ξ . In black, we show the result for f_2 using our new formalism, regulating the integral in Eq. (8.69) using a UV cutoff $\gamma_{UV} = 2\xi$. The gray-shaded band indicates the uncertainty of this result when increasing or decreasing γ_{UV} by a factor of three. In blue, we

⁷This numerical computation should not be confused with the one performed in the next section. Here, we refer to evaluating the integral in Eq. (8.74) by relying on the gauge-mode functions in Eq. (8.61). In the next section, we instead determine the mode functions numerically, without relying on the exact expressions in terms of Whittaker functions, which are only valid for constant ξ .

show the same result for the old source term used in Ref. [57], which neglects metric perturbations. The dashed blue line shows the result for $\gamma_{\text{UV}} = 2\xi$, and the light-blue shaded area shows the influence of varying this cutoff by a factor of three. Notably, for both the old source term and our new source term, the theory uncertainty from our regularization scheme is most pronounced for $\xi \lesssim 2$. It is also in this area, where our new source term deviates the most from the result obtained using the old source term. Finally, the orange dashed line shows the analytical result for $f_2(\xi)$ in Eq. (8.76) with $c_2 \approx 8.48 \times 10^{-5}$. Note that the analytical result only approximates the true numerical result even for large ξ , although the agreement is overall good.

The bispectrum

For the bispectrum, we again follow Ref. [57], and parametrize it in terms of a dimensionless function $f_3(\xi, x_2, x_3)$:

$$\mathcal{B}_\zeta(k_1, k_2, k_3) = \frac{3}{10} (2\pi)^{5/2} \left(\frac{H^2}{2\pi |\dot{\phi}_c|} \right)^6 \frac{e^{6\pi\xi}}{k^6} \frac{1 + x_2^3 + x_3^3}{x_2^3 x_3^3} f_3(\xi, x_2, x_3). \quad (8.77)$$

Here, we have defined $k = k_1$ as the largest of the three momenta k_1, k_2, k_3 , and the dimensionless variables $x_2 = k_2/k$ and $x_3 = k_3/k$ satisfy the inequalities in Eq. (8.48).

The function f_3 can be determined by substituting Eqs. (8.68) and (8.70) into Eq. (8.42), yielding

$$\begin{aligned} f_3(\xi, x_2, x_3) = & -\frac{5}{3\pi} \frac{e^{-6\pi\xi}}{x_2 x_3 (1 + x_2^3 + x_3^3)} \int d^3 \mathbf{p}_* \sum_{\lambda_1, \lambda_2, \lambda_3 = \pm 1} p_1 p_2 p_3 \\ & \times (\boldsymbol{\epsilon}_{\lambda_1}(\mathbf{p}_1) \cdot \boldsymbol{\epsilon}_{\lambda_2}^*(\mathbf{p}_2)) (\boldsymbol{\epsilon}_{\lambda_2}(\mathbf{p}_2) \cdot \boldsymbol{\epsilon}_{\lambda_3}^*(\mathbf{p}_3)) (\boldsymbol{\epsilon}_{\lambda_3}(\mathbf{p}_3) \cdot \boldsymbol{\epsilon}_{\lambda_1}^*(\mathbf{p}_1)) \\ & \times \mathcal{I}(\lambda_1, p_1; \lambda_2, p_2; 1) \mathcal{I}((\lambda_2, p_2)^*; \lambda_3, p_3; x_2) \mathcal{I}((\lambda_3, p_3)^*; (\lambda_1, p_1)^*; x_3), \end{aligned} \quad (8.78)$$

with $\mathbf{p}_1 = \mathbf{p}_*$, $\mathbf{p}_2 = \mathbf{p}_* - \hat{\mathbf{k}}_1$, and $\mathbf{p}_3 = \mathbf{p}_* + \hat{\mathbf{k}}_3 x_3$ for $\hat{\mathbf{k}}_i = \mathbf{k}_i/k_i$.

First, we study f_3 in the equilateral configuration with $x_2 = x_3 = 1$. Using the large- ξ approximation of the gauge-field mode functions, it scales as

$$f_3(\xi, 1, 1) \simeq \frac{c_3(1, 1)}{\xi^9}. \quad (8.79)$$

Evaluating Eq. (8.78) using our new source term, we find $c_3(1, 1) \approx -3.25 \times 10^{-7}$, while the old source term used in Ref. [57] yields $\tilde{c}_3(1, 1) \approx -2.77 \times 10^{-7}$, the former exceeding the latter by about 15%.⁸ Just as for the induced power spectrum, we also compute $f_3(\xi, 1, 1)$ numerically as a function of ξ . The results are shown in Fig. 8.2 using the same color scheme as in Fig. 8.1.

Next, we study the dependence for other momentum configurations. This is captured by the shape function Σ introduced in Eq. (8.45), which can be expressed in terms of f_3 as

$$\Sigma(\xi, x_2, x_3) = \frac{1 + x_2^3 + x_3^3}{3x_2 x_3} \frac{f_3(\xi, x_2, x_3)}{f_3(\xi, 1, 1)}. \quad (8.80)$$

We compute this shape function numerically for the fixed values $\xi = 1.5, 2, 2.5$, and 3 using $\gamma_{\text{UV}} = 2\xi$ for the UV cutoff. The results for this are shown as a density plot in Fig. 8.3, depicting

⁸Our results differ in sign with respect to the original result reported in Ref. [57]. This can be traced back to a difference in convention for the sign of ζ used in our work and the one in Refs. [56, 57], c.f. our Eq. (8.16) and Eq. (5.16) in Ref. [57].

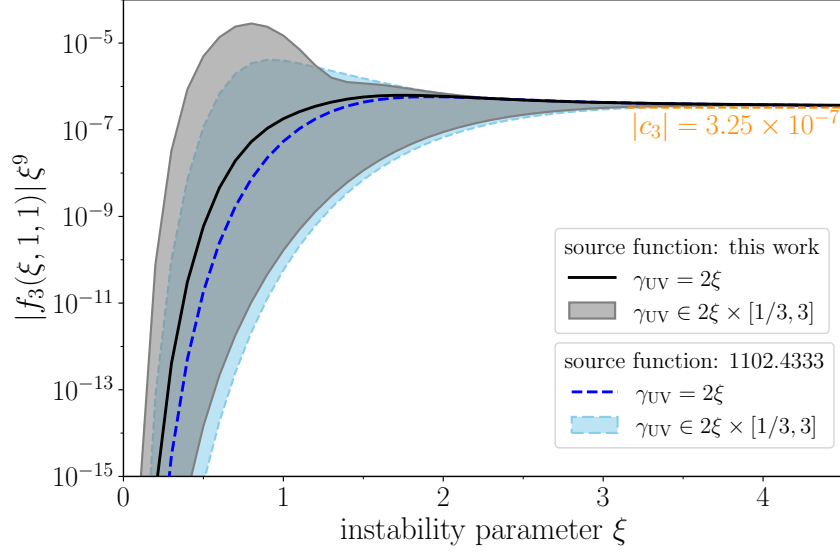


FIGURE 8.2: The function $f_3(\xi, 1, 1)$ parametrizing the equilateral scalar bispectrum multiplied by a factor ξ^9 . The colors and lines match the one for $f_2(\xi)$ given in Fig. 8.1. The orange dashed line shows the leading term in the expansion of $f_3(\xi, 1, 1)$ in the large- ξ approximation, see Eq. (8.79).

deviations from the equilateral configuration. For all values of ξ , we find that the gauge-field-induced bispectrum is broadly peaked around the equilateral configuration while being smallest for $x_3 \rightarrow 0$. This is also confirmed by table 8.1, wherein we list the cosines of the angles between $\Sigma(\xi, x_2, x_3)$ in Eq. (8.80) and the standard patterns in Eq. (8.46) using Eq. (8.49). The largest overlap is found for the equilateral configuration with the values for $\cos \Theta$ ranging from 0.94 to 0.97.

ξ	$\cos \Theta$ in Eq. (8.49) with respect to patterns:		
	Local (8.46a)	Equilateral (8.46b)	Orthogonal (8.46c)
1.5	0.47	0.97	-0.10
2.0	0.47	0.97	-0.10
2.5	0.47	0.94	-0.10
3.0	0.49	0.94	-0.10

TABLE 8.1: Cosines of the angles (8.49) between the shape functions $\Sigma(\xi, x_2, x_3)$ and the standard (local, equilateral, and orthogonal) patterns computed for $\xi = 1.5, 2, 2.5, 3$.

Based on this observation, we can then also confront our result for the bispectrum with CMB constraints on scalar non-Gaussianity.⁹ Data from the Planck satellite [15] constrains the amplitude of the bispectrum for the local, equilateral, and orthogonal patterns. In our case, the bispectrum is predominantly aligned with the equilateral configuration, for which the Planck constraints are given by $-73 < f_{\text{NL}}^{\text{equil}} < 21$ (68 % C.L.). Here, f_{NL} is a dimensionless parameter which quantifies the amplitude of the bispectrum. Following Ref. [57], it is related to $f_3(\xi, 1, 1)$ in Eq. (8.78) as

$$f_{\text{NL}}^{\text{equil}} = A_s f_3(\xi, 1, 1) e^{6\pi\xi}, \quad (8.81)$$

⁹Note that this comparison is not included in Ref. [2], but is based on the same results.

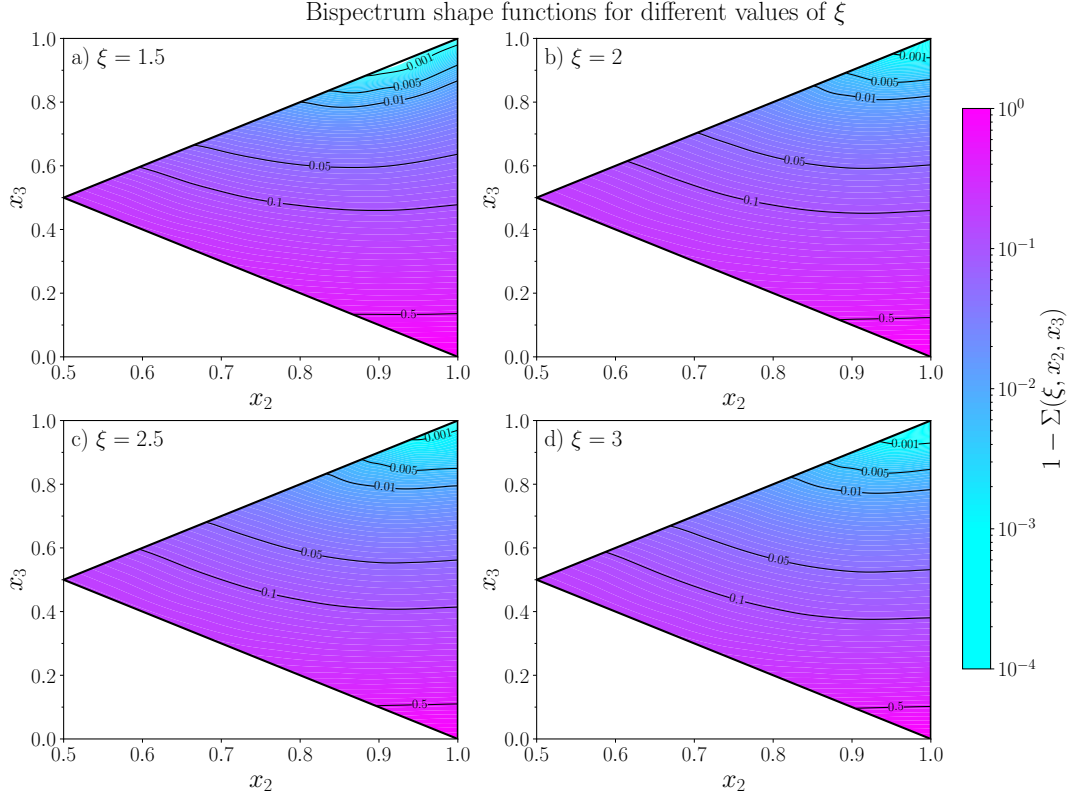


FIGURE 8.3: The bispectrum shape function $\Sigma(\xi, x_2, x_3)$ in the part of the (x_2, x_3) plane limited by the inequalities (8.48) for four fixed values of ξ : (a) $\xi = 1.5$, (b) $\xi = 2$, (c) $\xi = 2.5$, and (d) $\xi = 3$. Since in all cases the shape function exhibits a broad maximum at the equilateral configuration, the color scheme and contour labels correspond to the deviation of the shape function from unity, $1 - \Sigma(\xi, x_2, x_3)$ on a logarithmic scale.

with A_s the amplitude of the scalar power spectrum at CMB scales. Taking the observed value $A_s = 2.1 \times 10^{-9}$ [14], and the results for our numerical evaluation of $f_3(\xi, 1, 1)$, we find agreement with the known non-Gaussianity bound for axion inflation: $\xi \lesssim 2.5$ [65]. Note in particular, that this result appears to be insensitive to the uncertainties when varying the UV cutoff γ_{UV} . Although Fig 8.2 indicates a large sensitivity of $f_3(\xi, 1, 1)$ on the UV cutoff, the exponential dependence on ξ in Eq. (8.81) overshadows this uncertainty in the numerical prefactor. Furthermore, the result is not sensitive to the choice of source function, either the new one presented in this chapter, or the old one used in Ref. [57].

8.2.3 Discussion

Overall, we find a remarkable agreement between our comprehensive approach to computing scalar perturbations including the sourcing of inflaton and metric perturbations via the gauge field, compared to the treatment in Ref. [57], which only accounted for the sourcing of inflaton perturbations. This is most prominently seen in Fig. 8.1 and 8.2. In particular, we observe this agreement to be best in the regime $\xi \gtrsim 2$. Hence, we also find that our result confirms previous constraints on the degree of scalar non-Gaussianity for axion inflation, namely $\xi \lesssim 2.5$.

The fact that these two approaches to computing scalar perturbations agree so well for large ξ can be explained as follows. The important observation is that, for moderate to large ξ values, the dominant contribution to f_2 and f_3 comes from the positive-helicity gauge-field modes, as they

are exponentially amplified with respect to the negative ones (cf. Eq. (3.28)). Moreover, these enhanced modes are strongly peaked for momenta below the instability scale $2\xi aH$ (c.f. Fig. 3.1). These mode functions then enter in pairs into the integral $\mathcal{I}$ in Eq. (8.69), where the two functions share the largest overlap for near equal momenta, $p_* \simeq p'_*$, while quickly decreasing otherwise. The only difference between $\mathcal{I}$ for our new source term and the one used in Ref. [57] is given by the terms proportional to the functions h_1 , h_2 , and h_3 in Eq. (8.65). However, under the idealized conditions of constant ξ and H , all h_i 's are proportional to $(\lambda p_* - \lambda' p'_*) = (p_* - p'_*)$ for the case of two positive polarization modes, which vanish in the case of the largest overlap between mode functions, $p_* \simeq p'_*$. Thus, these new terms give a marginal contribution to $\mathcal{I}$, and therefore to the induced power spectrum and bispectrum, explaining the overall good agreement between our new treatment and the old one used in Ref. [57].

It should however be noted that these considerations only hold if we neglect gauge-field backreaction as we have done here. In particular, in the regime of strong backreaction, at least two of the above arguments are altered. As previously discussed in chapter 3, the slow-roll conditions are no longer satisfied in the case of strong backreaction, as both φ , $\langle E^2 \rangle$, $\langle B^2 \rangle$, etc. are subjected to violent oscillations. This means that additional terms in h_1 , h_2 , and h_3 could become important, which do not vanish in the case $p_* \simeq p'_*$. Moreover, since the oscillations in the inflaton and gauge field are also featured in the mode functions, their momentum distribution also becomes less peaked implying that contributions for $p_* \neq p'_*$ can become more relevant. In addition, gauge-field backreaction also modifies the left-hand side of the ζ equation, Eq. (8.17), as terms such as $\langle E^2 \rangle$ and $\langle B^2 \rangle$ can then contribute to the coefficients p_k and q_k in this differential equation. This directly alters the vacuum contribution, ζ^{vac} , but also affects the Green function which is needed to determine the induced contribution, ζ^{ind} . One can therefore expect that, in a scenario including gauge-field backreaction, the self-consistent inclusion of metric perturbations can have a more pronounced effect on the scalar power spectrum and bispectrum, which is why we investigate these effects in the next section in a fully numerical calculation of the scalar power spectrum, including strong-backreaction effects for both the inflationary background and the perturbations.

Before we discuss these numerical results, let us make some final remarks regarding the analytical calculation we have performed here. Firstly, it is also interesting to note that the importance of gauge-field-sourced metric perturbations appears to be more relevant in the case of small ξ . This is easy to understand following our previous line of reasoning. For smaller ξ , the overall importance of positive-helicity modes over negative ones becomes lessened. Thus, the terms proportional to the h_i 's in $\mathcal{I}$ contribute more, since their proportionality to $(\lambda p_* - \lambda' p'_*)$ does not entail a cancellation for unequal helicities even for $p_* \simeq p'_*$. Moreover, since gauge-field production is less efficient for smaller ξ , the positive-helicity modes are themselves also less peaked.

Meanwhile, it is important to note that these results for small ξ are overshadowed by a large theoretical uncertainty due to our regularization procedure. For a fixed value of ξ , the results for f_2 and f_3 can vary by up to six orders of magnitude for a mere $\mathcal{O}(1)$ change of the cutoff value γ_{UV} . Although the cutoff $\gamma_{\text{UV}} = 2\xi$ is well motivated, the sensitivity of our result to only small changes in this regulator indicates the necessity for a more rigorous and accurate regularization procedure based on conventional QFT techniques like adiabatic or dimensional regularization. To our knowledge, this has so far only been considered for gauge-field production in the absence of backreaction in Refs. [373, 374]. An application of proper renormalization techniques to the scalar power spectrum may be relevant for comparison with observations [444]. However, it should be noted that our simple estimates regarding the theoretical uncertainty of our regularization procedure did not affect the result for the non-Gaussianity bound, which stands at $\xi \lesssim 2.5$, irrespective of variations for γ_{UV} . Further investigations into these phenomenological implications of regularization techniques are nonetheless warranted, also regarding the bound on ξ for primordial-black-hole

overproduction in axion inflation, which also applies to the potentially sensitive region $\xi \simeq 2 - 2.5$. Recently, Ref. [445] took a first step toward investigating alternative regularization methods for the gauge-field-induced scalar power spectrum in the case of constant H and ξ . However, what is ultimately missing is the development and application of improved renormalization methods for the fully time-dependent case including strong gauge-field backreaction.

8.3 Second application: Axion inflation with backreaction

Having validated our new formalism in the weak-backreaction regime, where we could rely on (semi-)analytical techniques, we now seek to apply it to the study of scalar perturbations in situations where we account for the full complex dynamics of axion inflation. Note that the following discussion is based on ongoing work which is not part of our original publication, Ref. [2].

In our subsequent discussion, we merely focus on the calculation of the scalar power spectrum, not the bispectrum. At first, we discuss our general numerical strategy which we use to perform this computation. The basis for this is a similar calculation of the induced tensor power spectrum for axion inflation in Ref. [83], which we adapt to the study of scalar perturbations. We then apply this strategy by successively relaxing the assumptions from the previous section. In the first part, we compute the scalar perturbation spectrum on a fully dynamical background, but we set all backreaction terms ($\langle \mathbf{E}^2 \rangle$, $\langle \mathbf{B}^2 \rangle$ and $\langle \mathbf{E} \cdot \mathbf{B} \rangle$ etc.) explicitly to zero when solving the ζ -equation, Eq. (8.17). This allows us to directly compare our numerical results to our earlier, semi-analytical ones. Only after we have performed this validation do we study the situation where the backreaction terms entering Eq. (8.17) are directly taken from the numerical background. In what follows, we shall often refer to these two assumptions as “neglecting backreaction terms” and “including backreaction terms”. We want to state explicitly that these two statements only refer to the calculation of the scalar power spectrum. The background dynamics are always solved in their entirety, including gauge-field backreaction terms, as explained in chapter 4.

8.3.1 Computing the scalar power spectrum using the GEF

The overall strategy to numerically compute the scalar power spectrum remains the same as in the previous section. The scalar power spectrum still consists of two separate contributions, the vacuum power spectrum, $\mathcal{P}_\zeta^{\text{vac}}$, given in Eq. (8.22), and the induced power spectrum, $\mathcal{P}_\zeta^{\text{ind}}$, given in Eq. (8.30). To compute these contributions, we therefore need to determine the vacuum mode functions $u_\lambda(\eta, k)$, the Green function $G_k(\eta, \tau)$, and the source term $K_{\lambda, \mathbf{p}; \lambda', \mathbf{k}-\mathbf{p}; k}(\tau)$. The difference to the previous section is that we now determine these quantities for a fully dynamical background.

Background dynamics

In a first step, we need to solve the dynamical background equations for axion inflation. For this, we rely on the GEF approach outlined in chapter 4, which gives us access to the full time evolution of the background inflaton field, the Hubble rate, the backreaction parameters $\langle \mathbf{E}^2 \rangle$, $\langle \mathbf{B}^2 \rangle$, $\langle \mathbf{E} \cdot \mathbf{B} \rangle$, and so on. Moreover, the GEF also allows us to determine the spectrum of gauge-field mode functions $A_\lambda(t, k)$, which enters into the source function for the induced scalar power spectrum in Eq. (8.39b).

In what follows, we consider the situation where the GEF equations are initialized with slow-roll initial conditions. This means, we assume gauge-field production is initially absent in our system, and is free to grow as time progresses. This corresponds to the slow-roll initial conditions

for the GEF that we discussed in Sec. 4.2.2. Specifically, in what follows, we compute background dynamics for our standard benchmark scenario: a quadratic inflaton potential, $V(\phi) = m^2\phi^2/2$ with mass $m = 6.16 \times 10^{-6} M_{\text{P}}$, and the inflaton field initialized on the slow-roll trajectory such that we are guaranteed at least 60 e -folds of inflation.

Vacuum power spectrum

The vacuum contribution $\mathcal{P}_\zeta^{\text{vac}}$ can be computed from the mode functions $u(\eta, k)$ of ζ^{vac} using Eq. (8.39a). These mode functions need to be determined numerically by integrating the mode equation

$$u''(\eta, k) + p_k(\eta)u'(\eta, k) + q_k(\eta)u(\eta, k) = 0. \quad (8.82)$$

Here, the coefficients p_k and q_k are now both fully time-dependent, and follow from the general expressions given in App. D. They depend on wavenumber k , and the background dynamics, and can therefore be derived from the GEF solution. To solve this mode equation, we apply the strategy detailed in App. B.

What requires some care is the initialization of these mode functions. When neglecting backreaction, we have that $p_k = 2(z'/z)$ with $z = a\dot{\phi}/H$ and $q_k = k^2$ as in Sec. 8.2. The mode functions $u(\eta, k)$ can therefore be initialized in the standard Bunch–Davies vacuum, i.e., $v(\eta, k) \equiv z(\eta)u(\eta, k) \simeq e^{-ik\eta}/\sqrt{2k}$ for $|k\eta| \gg 1$ (see Eq. (8.54)).

Less obvious is the initialization for the case when backreaction terms, $\langle E^2 \rangle$, $\langle B^2 \rangle$ etc., are included in p_k and q_k . In this case, we can however take advantage of the fact that we initialize our GEF system assuming slow-roll initial conditions. This means, at early times in our system's evolution, gauge-field backreaction is not yet effective. In particular, backreaction in the Friedmann equation is absent, entailing $6M_{\text{P}}^2 H^2 \gg \langle E^2 + B^2 \rangle$ (see also Eq. (3.27)). For a mode $u(\eta, k)$ deep inside the horizon, $k \gg aH$, we then also have that $k^2 \gg a^2 \langle E^2 + B^2 \rangle / M_{\text{P}}^2$. In this limit, the expressions for p_k and q_k again reduce to $p_k \simeq 2(z'/z)$ and $q_k \simeq k^2$ as in the case of no backreaction. The mode functions $u(\eta, k)$ can therefore be initialized as before;

$$u(\eta, k) \simeq \frac{1}{z(\eta)} \frac{1}{\sqrt{2k}} e^{-ik\eta}, \quad \text{for } |k\eta| \gg 1 \text{ and } H^2 \gg \frac{\langle E^2 + B^2 \rangle}{6M_{\text{P}}^2}. \quad (8.83)$$

Induced power spectrum

To compute the induced power spectrum, $\mathcal{P}_\zeta^{\text{ind}}$, one first needs to determine the Green function $G_k(\eta, \tau)$. One may naively expect that the Green function for fixed k can be simply computed based on the numerical mode function $u(\eta, k)$ and its complex conjugate as in Eq. (8.27). In practice, however, this approach yields numerically unstable results since the mode function $u(\eta, k)$ are highly oscillatory at early times resulting in undesirable cancellation effects. An improved numerical approach entails computing the $G_k(\eta, \tau)$ for a fixed τ and k by directly evolving it in time. Details on this method are given in App. B.¹⁰

Alongside the Green function, the induced power spectrum also requires the determination of the source function $K_{\lambda, p; \lambda', p'; k}(\eta)$, the general expression for which is given in App. D. As previously explained, it decomposes into five terms, each of which is proportional to one of the coefficients $r^{(q)}$, $r^{(\rho)}$, $r_1^{(\sigma)}$, $r_2^{(\sigma)}$ and $r^{(E \cdot B)}$ appearing in Eq. (8.19). The lengthy expressions for these coefficients are also relayed to App. D, but for our subsequent discussion, it suffices to know that they depend on wavenumber k and background quantities like φ , H , $\langle E \rangle^2$, etc. They can therefore be readily computed from the background GEF evolution. Finally, the source term

¹⁰Note that this strategy has not been used in Ref. [83], who relied on an analytical approximation for the Green function in their case.

$K_{\lambda,p;\lambda',p';k}(\eta)$ also depends on the gauge-mode functions $A_\lambda(\eta, k)$, which, as explained above, are also determined as part of the GEF technique (see chapter 4).

Finally, $\mathcal{P}_\zeta^{\text{ind}}(\eta, k)$ can be computed by numerically evaluating both the momentum and time integral in Eq. (8.39b). These integrals are successively calculated for a discrete set of momenta and for a fixed set of times. For the regularization of this integral, we rely on the standard momentum cutoff at the instability scale k_h given in Eq. (3.24). Once again, we refer the reader to App. B for details on this numerical integration strategy.

8.3.2 Neglecting gauge-field backreaction

For a controlled first application and validation of our numerical strategy, we now apply it to the scenario where all backreaction terms $\langle \mathbf{E} \rangle^2$, $\langle \mathbf{B}^2 \rangle$, $\langle \mathbf{E} \cdot \mathbf{B} \rangle$ etc., are set explicitly to zero when computing the coefficients p_k , and q_k , and the r 's entering in Eq. (8.82) and the source term $K_{\lambda,p;\lambda',k-p;k}(\eta)$.

Validation: $\beta = 10$ — weak backreaction dynamics

As a first benchmark, we compute the vacuum and induced scalar power spectrum for the case of a weak axion–vector coupling, $\beta = 10$. As in Sec. 8.2, we compute the induced scalar power spectrum once for the source term K that we derive in this chapter, i.e., the one that accounts for the sourcing of metric perturbations via the gauge field, and once for the original source term $\tilde{K}$ derived in Ref. [57]. For the scenario we are considering here, $\beta = 10$, gauge-field production is not efficient enough for the background evolution to be subjected to gauge-field backreaction effects, and the end of inflation is reached on the slow-roll attractor. Since the background evolution is given entirely by slow-roll dynamics, we are therefore in a scenario where our analytical results for $\mathcal{P}_\zeta^{\text{vac}}$ and $\mathcal{P}_\zeta^{\text{ind}}$ from the previous section (Eqs. (8.72) and (8.73)) should generally agree with our numerical ones. Moreover, based on these analytical results, we can expect good agreement between the induced power spectrum computed for the two source functions, K and $\tilde{K}$. This first benchmark therefore mainly serves as a validation of the integrity of our numerical calculation.

In Fig. 8.4, we show our results for this benchmark scenario. The upper panel depicts the vacuum power spectrum, while the lower panel gives the induced power spectrum, both of which are computed at the end of inflation. We show the results for numerical calculations in red and in blue, respectively for our new source function including sourced metric perturbations, and the original source function used in Ref. [57]. For comparison, the dotted black lines indicate the analytical results from the previous section. These analytical results are obtained by evaluating Eqs. (8.72) and (8.73) based on the time-dependent values of ξ , H , etc. and we approximate $f_2(\xi)$ assuming the large- ξ approximation, Eq. (8.76) with $c_2 = 8.48 \times 10^{-5}$.

For the vacuum power spectrum, we can note excellent agreement between the numerical approach and the analytical results. For the induced power spectrum, the agreement between both numerical results is near perfect, precisely as expected from our findings of the previous section. Meanwhile, the analytical approximation only qualitatively reproduces the numerical results, with deviations that are especially pronounced at low wavenumbers, corresponding to modes that exit the horizon at early times. Such differences for $\mathcal{P}_\zeta^{\text{ind}}$ are however expected, seeing as the analytical result assumes large ξ values, which are not realized for the dynamical system we consider here; for $\beta = 10$, ξ remains at $\xi \lesssim 2$ until the late stages of the inflationary evolution. As seen in Fig. 8.1, $f_2(\xi)$ can deviate significantly from the value in the large- ξ approximation.

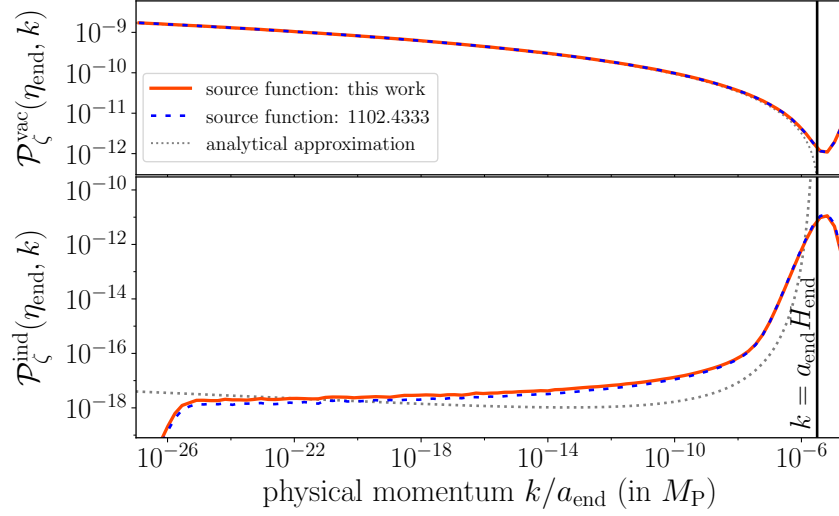


FIGURE 8.4: The vacuum (upper panel) and induced (lower panel) scalar power spectrum at the end of inflation for a dimensionless axion–vector coupling $\beta = 10$. In red, we show results for the numerical computation of the power spectra using the source term K derived in this chapter. Here, we are setting all gauge-field backreaction terms ($\langle \mathbf{E} \rangle^2$, $\langle \mathbf{B}^2 \rangle$, $\langle \mathbf{E} \cdot \mathbf{B} \rangle$ etc.) entering Eq. (8.17) to zero. In dashed blue, we show numerical results for the power spectra which we compute for the original source term presented in Ref. [57]. The dotted black line shows analytical results for both the vacuum and induced power spectra based on Eqs. (8.72) and (8.73) with $\mathcal{P}_\zeta^{\text{ind}}$ computed assuming the large- ξ approximation, Eq. (8.76).

Extrapolation: $\beta = 25$ — from weak to strong-backreaction dynamics

After this first consistency check, we extend our calculation to a benchmark scenario with a larger value for the axion–vector coupling, $\beta = 25$. As a reminder, we discussed the background dynamics for this scenario in Sec. 4.2.2, finding that the system enters the regime of strong gauge-field backreaction, which includes a prolonged duration of inflation and a phase of violent oscillations in the inflaton velocity (and therefore of ξ). Unlike the previous benchmark point, we can here consider how strong-backreaction effects can affect the scalar power spectrum. Specifically, as we are considering slow-roll initial conditions, the scalar power spectrum should qualitatively separate into two distinctive regimes: the power spectrum at low wavenumber reflects scalar modes that have exited the horizon during a stage of slow-roll inflation, while larger wavenumbers correspond to modes that exited the horizon during a stage of strong backreaction. Once again, we compute the vacuum and the induced power spectrum for this benchmark point, once based on the analytical formula in Eqs. (8.72) and (8.73), once numerically for the new source term derived in this chapter, and then again for the old source term used in Ref. [57]. The results of this computation are then given in Fig. 8.5 with the same colors as in Fig. 8.4.

Looking first at the analytical results (black dotted) in Fig 8.5, we can recognize the separation of the power spectrum into the low-wavenumber part corresponding to weak-backreaction dynamics, and the large-wavenumber section corresponding to modes that have exited the horizon in the strong-backreaction regime. For the analytical results, the curves we are plotting for $\mathcal{P}_\zeta^{\text{vac}}$ and $\mathcal{P}_\zeta^{\text{ind}}$ are direct functions of the inflaton velocity $\dot{\varphi}$ (or the instability parameter, ξ). Therefore, we can identify the onset of strong backreaction dynamics in these power spectra via the violent oscillations in the dotted black curves.

Let us first focus on the results at low wavenumbers. In this case, the analytical and numerical results are all in good agreement, which is exactly what we should expect. Firstly, analytical and

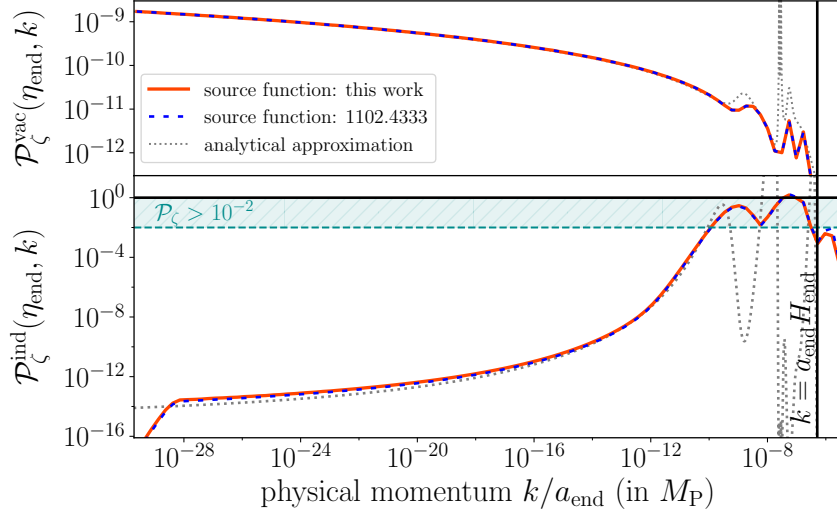


FIGURE 8.5: Equivalent results as in Fig. 8.4, this time for $\beta = 25$. Colors and line styles match those in Fig. 8.4. We also highlight the regime where the induced scalar power spectrum surpasses the threshold $\mathcal{P}_\zeta^{\text{ind}} > 10^{-2}$, indicating the potential for PBH formation.

numerical results for $\mathcal{P}_\zeta^{\text{vac}}$ and $\mathcal{P}_\zeta^{\text{ind}}$ should agree, as they are computed for approximate slow-roll dynamics, and, in particular the large- ξ approximation for the analytical $\mathcal{P}_\zeta^{\text{ind}}$ should perform better than for $\beta = 10$, as the larger value of β that we consider here implies a larger value of ξ already at early times during inflation. Finally, the agreement between the two numerical evaluations of $\mathcal{P}_\zeta^{\text{ind}}$ for low wavenumbers is entirely consistent with our findings in Sec. 8.2. Overall, this agreement, as well as the agreement that we found for $\beta = 10$ validates the consistency of our numerical implementation.

Next, let us consider results at large wavenumbers, after slow-roll dynamics have broken down. In this case, we can observe substantial differences between the analytical and the numerical results both for the induced and the vacuum power spectrum. These deviations indicate the breakdown of the analytical calculation, not the numerical one, as the analytical results are evaluated in a dynamical regime where the underlying assumptions under which they have been computed are no longer valid. Instead, we can focus on comparing the two numerical results for the induced power spectra (lower panel) to assess the importance of gauge-field-induced metric perturbations. We can note that the result for $\mathcal{P}_\zeta^{\text{ind}}$ for our new source term derived in this chapter, and the original source term from Ref. [57], agree well, also for strong-backreaction effects. This indicates that gauge-field-induced metric perturbations by themselves give a subdominant contribution to the induced power spectrum, and that the dominant contribution has already been accounted for in the source term derived in Ref. [57].

At this stage, we have however not taken into consideration the second effect which our formalism can account for, which is the effect of the backreaction terms $\langle \mathbf{E}^2 \rangle$, $\langle \mathbf{E} \cdot \mathbf{B} \rangle$ etc. on the evolution of curvature perturbations. We shall consider this in the next section.

Before we do so however, we should briefly pause, as the results in Fig. 8.5 also serve as an illustration why a reliable computation of the induced scalar power spectrum in the strong-backreaction regime is relevant. In Fig. 8.5, we indicate that the induced power spectrum surpasses the threshold $\mathcal{P}_\zeta^{\text{ind}} > 10^{-2}$ for large wavenumbers. As mention in chapter 1, this threshold is conventionally taken as a benchmark value that indicates the potential for primordial-black-hole (PBH) formation once scalar perturbations of the corresponding wavenumbers reenter the horizon after inflation [253]. The fact that $\mathcal{P}_\zeta^{\text{ind}}$ surpasses this value in the strong-backreaction

regime signals that axion inflation may indeed lead to the formation of PBHs. Meanwhile, this demonstrates that accurate calculations of their formation rate should take strong-backreaction effects into account, as analytical results are no longer reliable in this dynamical regime, as had also been pointed out in Ref. [109]. At the same time, the large amplitude of $\mathcal{P}_\zeta^{\text{ind}}$ may also signal a breakdown of perturbativity during the strong-backreaction regime, which would align with the observation made using lattice techniques in, e.g., Refs. [378, 379] (see also our discussion in chapter 4). In either case, this emphasizes the need for an accurate determination of the induced power spectrum when including strong-backreaction effects, and highlights the importance of the GEF-based strategy we have developed and applied in this section.

8.3.3 Including gauge-field backreaction

Let us then consider how the results presented in the previous section change, when we include backreaction terms into the calculation of the scalar power spectrum. To once again make the distinction explicit, in this section, “no backreaction” refers to setting $\langle \mathbf{E} \rangle^2 = \langle \mathbf{B}^2 \rangle = \langle \mathbf{E} \cdot \mathbf{B} \rangle = 0$ for the computation of the scalar power spectrum, while “including backreaction” implies these terms are consistently accounted for in the calculation of $\mathcal{P}_\zeta^{\text{ind}}$. In the latter case, the values of the backreaction terms are directly taken from the GEF solution for the dynamical background. Moreover, we are here entirely reliant on the source function K derived in this chapter, as the one presented in Ref. [57] did not account for backreaction effects.

As before, the first thing we want to consider is the application to the case $\beta = 10$. Here, we expect to find that the backreaction terms $\langle \mathbf{E} \rangle^2$, $\langle \mathbf{B}^2 \rangle$, $\langle \mathbf{E} \cdot \mathbf{B} \rangle$ only marginally affect the results for the overall scalar power spectrum, since this scenario features weak gauge-field production, and the backreaction terms remain small (relative to, e.g., the Hubble rate, H).

In Fig. 8.6, we show first results for the induced scalar power spectrum in this scenario. Here, we compare the numerical results for $\mathcal{P}_\zeta^{\text{ind}}$ without backreaction terms (gray, solid), to the result when backreaction terms are included (shades of blue, dotted). The spectrum for the no-backreaction case is computed at the end of inflation, while the results including backreaction terms are given at 0, 3, 6 and 9 e -folds before the end of inflation as indicated by the colorbar on the right of the figure. At 9 e -folds before the end of inflation, the power spectrum including backreaction terms matches our expectation; the results with and without backreaction terms agree for all modes that have exited the horizon at this moment of time. The results for the power spectrum at later times continue to align with the no-backreaction result for large wavenumbers. However, at low wavenumbers, there is a striking deviation. Small wavenumbers that have exited the horizon many e -folds ago, and for which the scalar power spectrum should remain frozen, instead undergo rapid growth. The induced scalar power spectrum develops the behavior $\mathcal{P}_\zeta^{\text{ind}} \propto 1/k$ for these modes, indicating an infrared divergence which is not present for the scalar power spectrum where backreaction terms are neglected.

Mathematically, this effect can be understood by further analyzing the source term for $\mathcal{P}_\zeta^{\text{ind}}$. As previously noted, it is composed of five terms, each proportional to one of the r -coefficients in Eq. (8.19). By studying numerical results for these coefficients both with and without including backreaction terms, one can identify particularly strong effects of these backreaction terms for the coefficient $r_1^{(\sigma)}$. To illustrate this, we show the time evolution of this coefficient for the case $\beta = 10$ in Fig. 8.7. Here, we compute $r_1^{(\sigma)}$ once without backreaction terms (dashed lines), and once when including them (solid lines). We do this for three momenta $k_1 < k_2 < k_3$. What is immediately evident, is that $|r_1^{(\sigma)}|$ is approximately time-independent at early times, which is reproduced both with and without including backreaction terms into its determination. At late times, this remains true for the computation without backreaction terms, while $|r_1^{(\sigma)}|$ with backreaction terms starts

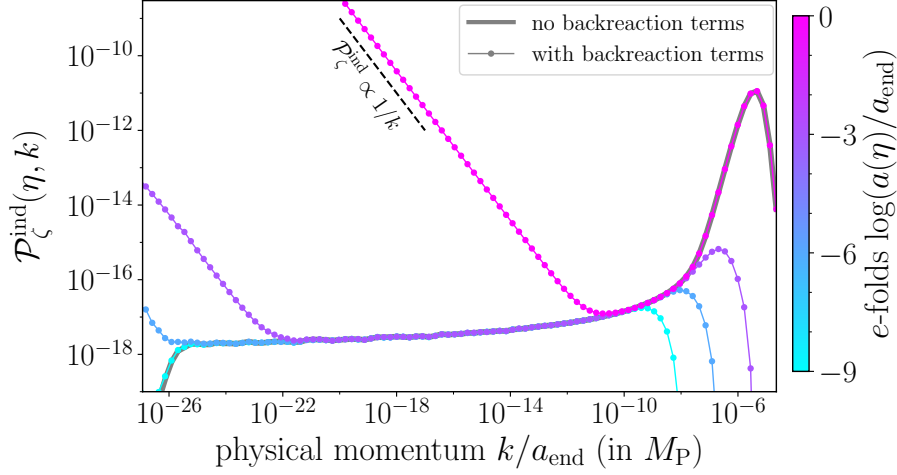


FIGURE 8.6: Results for the induced scalar power spectrum once backreaction terms ($\langle \mathbf{E} \rangle^2$, $\langle \mathbf{B}^2 \rangle$, $\langle \mathbf{E} \cdot \mathbf{B} \rangle$ etc.) are included into the determination of ζ . We show four realizations of the $\mathcal{P}_\zeta^{\text{ind}}$ computed at four different times during inflation; 9, 6, 3 and 0 e -foldings before the end of inflation. The colors of the curves correspond to the scale given on the right-hand side of the plot. Underlying these curves in gray, we show the result for $\mathcal{P}_\zeta^{\text{ind}}$ at the end of inflation when backreaction terms are not included in the determination of ζ . The gray curve in this figure matches the one shown in red in Fig. 8.4. Clearly visible is a growing infrared divergence in the induced scalar power spectrum, which develops when backreaction terms are included in its determination.

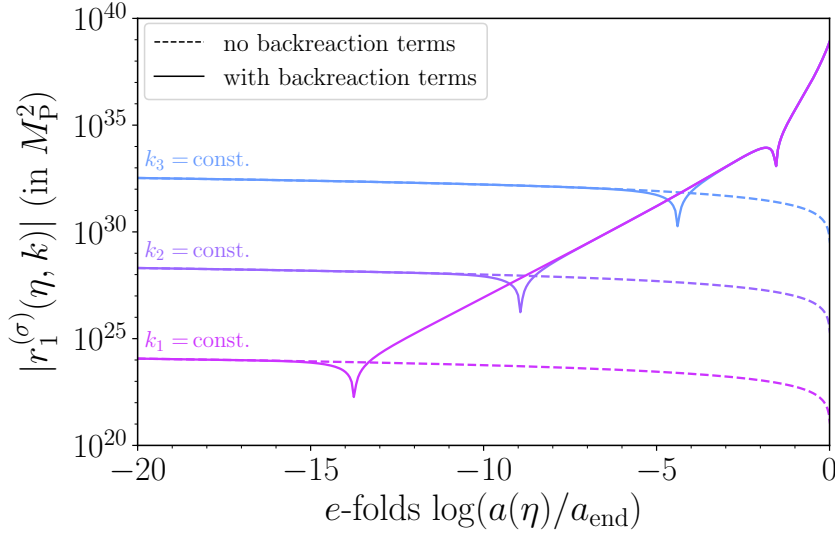


FIGURE 8.7: The time evolution of the coefficient $r_1^{(\sigma)}$ which enters into the computation of $\mathcal{P}_\zeta^{\text{ind}}$. The coefficient is computed for three wavenumbers $k_1 < k_2 < k_3$, once without the inclusion of backreaction terms (dashed), and once when backreaction terms are included (solid). The late-time growth of $|r_1^{(\sigma)}|$ in the latter case can be attributed to the IR divergence observed in Fig. 8.6.

suddenly growing. This growth affects all wavenumbers, first for the smallest wavenumber, k_1 , and successively later for k_2 and k_3 .

A thorough investigation of the expression for $|r_1^{(\sigma)}|$ in Eq. (D.18e) reveals that it indeed contains terms proportional to, e.g., the combination $k^2 + a^2 \mathcal{X}/M_{\text{P}}^2$ with $\mathcal{X} = \langle \mathbf{E}^2 \rangle, \langle \mathbf{B}^2 \rangle$. As gauge-field production continuously grows during inflation, the contribution from $\langle \mathbf{E}^2 \rangle, \langle \mathbf{B}^2 \rangle$ thus begins to take over at late times, first for smaller values of k , then for larger ones. This growth is evidently not present when explicitly setting $\langle \mathbf{E}^2 \rangle = \langle \mathbf{B}^2 \rangle = 0$, as we did in the calculations of the power spectrum in the previous sections. Because $r_1^{(\sigma)}$ immediately enters the source term of $\mathcal{P}_\zeta^{\text{ind}}$ (see Eq. (D.20c)), its growth can then be directly attributed to the development of the infrared divergence in the induced scalar power spectrum that we can see in Fig. 8.6; starting with the smallest wavenumbers, the source contribution proportional to $r_1^{(\sigma)}$ begins to dominate the spectrum, progressively affecting larger and larger modes as inflation proceeds. Indeed, we confirmed that the result for $\mathcal{P}_\zeta^{\text{ind}}$ does not exhibit the infrared divergence when backreaction effects are only discounted for $r_1^{(\sigma)}$, but kept for all other r -coefficients as well as in p_k and q_k .

Although this mathematical explanation appears sound, it does not shed any light on the physical origin of the IR divergence we observe. The separation into five contributions is done for mathematical convenience, but holds no deeper physical meaning. Thus, identifying the instigating coefficient $r_1^{(\sigma)}$ does not allow us to gain insight into a possible physical origin of the IR divergence of $\mathcal{P}_\zeta^{\text{ind}}$ that we observe. At this point, it merely supports the interpretation that this effect is not due to an inconsistency in our numerical implementation, but rather has a clear mathematical origin.

The IR divergence is clearly the cause for further investigation, as it is a significant deviation from our expectation that, for $\beta = 10$, the effect of the backreaction terms $\langle \mathbf{E}^2 \rangle, \langle \mathbf{B}^2 \rangle, \langle \mathbf{E} \cdot \mathbf{B} \rangle$ etc., on the scalar power spectrum should not be large. Moreover, we find that similar features appear also for larger β values. This raises the question if this result is physically meaningful, or if it is rooted in one of our underlying assumptions when developing the formalism we present in this chapter. At present, we are not able to draw a definitive conclusion regarding this question. Instead, in what follows, we briefly highlight some observations and deliberations on this matter.

Deliberations on the infrared divergence

In assessing the importance of the IR-divergent scalar power spectrum, two stances appear defensible from our perspective. The first is to identify the sudden growth in the scalar power spectrum as a real physical effect. In this case, the IR-divergent spectrum signals a clear breakdown of the perturbative approach that we relied on in this chapter. The second is that the IR divergence can be attributed to an effect which our formalism does not take into account.

Let us first direct our attention toward the assumption of perturbativity. Here, some insight can be gained by considering the results of lattice simulations for axion inflation, e.g., Refs. [378–380]. As we have noted in chapter 4, these simulations indicate that gauge-field production in the regime of strong backreaction leads to a significant growth of inflaton perturbations, which then also affect gauge-field production. In principle, this dynamical backreaction of inflaton inhomogeneities could imply that a separation into background and perturbations as we have considered here is not valid, and that the IR divergence we see in Fig. 8.6 is an indication of this breakdown of perturbativity.

However, this seems unlikely, since, for the benchmark point we have analyzed in Fig. 8.4, we are studying a scenario of weak gauge-field production for which lattice simulations indicate that these inhomogeneous-backreaction effects are irrelevant. One caveat to this conclusion is that current lattice simulations assume an isotropic and homogeneous spacetime metric, i.e., they do not account for the sourcing of metric perturbations via the gauge field. In principle, one could therefore suppose that the breakdown of perturbativity comes from the fact that metric perturbations

are effectively sourced by gauge fields, and that metric perturbations themselves are an effective source of gauge fields, purely via a gravitational interaction. However, given that our results in the previous sections have consistently demonstrated that the contribution from gauge-field-induced metric perturbations is subdominant with respect to sourced inflaton perturbations, it is unclear why the reversed backreaction of metric perturbations on gauge-field production should be more important than that of inflaton perturbations on gauge-field production, which we know to be irrelevant for $\beta = 10$. Overall, it seems unlikely that the IR divergence we observe in Fig. 8.6 is indeed signalling a breakdown of perturbation theory.

Let us then consider an alternative origin for the observed IR divergence. Here, we may speculate on the possibility that our formalism requires further refinement, with the IR divergence being a symptom of an inconsistency in our treatment. At present, we perceive two possibilities in this regard.

The first possibility relates to the difference between adiabatic and isocurvature perturbations. As we have noted in chapter 1, single-field slow-roll inflation leads to the production of adiabatic perturbations, for which one generically finds that the super-horizon spectrum of curvature perturbations is conserved (see, e.g., the standard discussion in textbooks on this matter [138, 191]). The presence of gauge fields adds an isocurvature contribution to the overall spectrum of scalar perturbations. This entails that the scalar power spectrum need not be frozen on super-horizon scales, which may lead to a large-scale growth in $\mathcal{P}_\zeta^{\text{ind}}$ as we observe it. An interesting thing to note here is that studies of isocurvature perturbations for multifield inflation scenarios have indeed identified similar IR divergences in the scalar power spectrum (see, for instance, Ref. [446]). The correct approach to studying isocurvature perturbations in multifield inflation scenarios is to then collectively solve a set of perturbation equations for all fields in the system (see Ref. [155, 434, 435]). As we have remarked upon in Sec. 8.1.2, the analogous treatment of isocurvature perturbations in our model would be to solve our master equation for ζ in Eq. (8.17) together with the equations of motion for the source terms F_α in Eq. (8.14). This approach to studying scalar perturbations in axion inflation entails a significant departure from the approach we have presented here, and would require a thorough reexamination of our formalism.

A second possibility is that the IR divergence we encounter in our model is unrelated to isocurvature perturbations. The basis for this speculation is a recent discussion concerning the production of scalar perturbations for ultra-slow-roll (USR) models of inflation [447–449]. In models of USR inflation, one considers a scenario of single-field inflation, where the inflaton undergoes a transition through a nearly flat part of the inflaton potential. In this flat part, the inflaton velocity becomes small, which amplifies the standard vacuum power spectrum for inflation (see Eq. (8.72), in particular the proportionality to $1/\dot{\varphi}^2$.) The authors of Ref. [447] studied the contribution to the scalar power spectrum sourced by inflaton self-interactions. They observed a similar large-scale enhancement of the scalar power spectrum as the one we find for axion inflation, albeit in a single-field inflation scenario, where isocurvature perturbations are absent. In Ref. [447], it was concluded that these effects signal a breakdown of the perturbative description of USR inflation, while other authors refuted these claims [448]. Eventually, Ref. [449] proved that these effects are spurious. To arrive at this conclusion, Ref. [449] worked in a suitably chosen gauge (spatially-flat gauge), which simplified the complexity of their computation. Although our Eq. (8.17) is gauge-invariant, it may similarly be the case that a rederivation of this equation in spatially-flat gauge reveals inconsistencies in our previous derivation, which may resolve the IR divergence we observe in our present analysis.

Ultimately, seeking the answer to these questions is beyond the scope of this thesis. However, seeing as the numerical tools we have developed in this section revealed the IR divergent spectrum, they may also prove essential in gaining further insights into answering them in the future.

8.4 Summary and conclusion

In this chapter, we have presented a new formalism to study the production of scalar perturbations during axion inflation. This formalism entails a self-consistent application of cosmological perturbation theory, and accounts for the creation of scalar inflaton and metric perturbations via the inflationary gauge field present during axion inflation. In this formalism, scalar perturbations can be directly sourced by an inhomogeneous gauge-field source, while the presence of a non-trivial gauge-field background can also backreact on the evolution of scalar perturbations. As long as these backreaction effects are neglected, we find the contribution of gauge-field-sourced metric perturbations to be small compared to sourced inflaton perturbations. Meanwhile, including gauge-field-backreaction effects leads to the development of an infrared (IR) divergence in the spectrum of scalar perturbations. At present, it is unclear if this second effect is physical in nature, signalling the breakdown of perturbativity, or if the divergence requires a more refined version of the formalism we have presented here.

These conclusions are based on a comprehensive application of our formalism in two case studies. In the first case study, we relied on a fully analytical study of scalar perturbations, where we assumed that gauge-field backreaction is irrelevant both when studying the inflationary background, and for the creation of scalar perturbations. By assuming idealized slow-roll conditions, a constant Hubble rate, H , and a constant inflaton velocity, $\dot{\phi}$, we were able to perform a comprehensive comparison between our new formalism and prior literature results [57]. This comparison suggests that gauge-field-sourced metric perturbations only marginally affect the amplitude of scalar perturbations if gauge-field production is strong, as quantified by large values of the instability parameter, $\xi \gtrsim 3$. Although we have observed a more pronounced effect of gauge-field-induced metric perturbations for smaller ξ , we have however found that their additional contribution does not appear to be phenomenologically relevant for axion inflation: we find that the difference in amplitude for the scalar bispectrum at $\xi \lesssim 3$ is not significant enough to affect the existing CMB bound on the degree of scalar non-Gaussianity in axion inflation, which remains at $\xi \lesssim 2.5$.

One possibility would then be that gauge-field-sourced inflaton perturbations are more relevant in the case where gauge-field backreaction has effectively disrupted the slow-roll evolution of the inflaton, i.e., in the strong-backreaction regime. To assess this possibility, we developed and applied a fully numerical technique to compute the scalar power spectrum, using the dynamical information on the inflationary background from the gradient-expansion formalism (GEF) as a basis for this computation. After careful calibration, the application of this technique to a benchmark model featuring strong gauge-field backreaction on the inflaton's motion however still indicates that metric perturbations remain a subdominant contribution to the total scalar power spectrum, at least for as long as gauge-field backreaction on the evolution of scalar perturbations is itself neglected. Instead, we find that the inclusion of backreaction effects leads to the build-up of large-scale scalar perturbations over time, resulting in the aforementioned IR divergence in the scalar power spectrum.

One interpretation of this IR divergence is that it signals a breakdown of perturbativity. However, this appears to be inconsistent with prior lattice studies of axion inflation (e.g., Refs. [378, 379]), as the infrared divergence occurs in a dynamical regime for which lattice studies find a negligible scalar backreaction on gauge-field production. However, these lattice studies do not account for the inhomogeneous evolution of the spacetime metric, which would require a full application of numerical relativity. We can therefore not draw a definitive conclusion regarding the validity of our perturbative assumption, but, given that our other results indicate a subdominant effect of metric perturbations, it appears unlikely that their backreaction may drastically affect gauge-field production.

Meanwhile, this backreaction of metric perturbations is precisely neglected in our present analysis. Going beyond this assumption requires to consistently solve the evolution equation for gauge-field perturbations alongside that of scalar perturbations. This is analogous to the study of multifield models of inflation [155, 434, 435], where similar infrared divergences are otherwise known to occur [446]. Another possibility is that the divergent effect we find for axion inflation is comparable to those found in models of ultra-slow-roll inflation [447]. These effects have recently been demonstrated to be spurious [450] by working in a suitably chosen gauge which simplified the complexity of the calculation. Similar considerations may apply to our study of axion inflation. Ultimately, understanding the origin of the IR divergence in the scalar power spectrum is paramount for the success of our novel formalism, but lies outside the scope of the present thesis.

Irrespective of the open questions regarding IR divergences, the results we have presented in this chapter highlight the necessity for pursuing a more refined strategy for computing scalar perturbations as we have done it here. Even when disregarding the effects of gauge-field backreaction for the study of scalar perturbations, we observe that late-time gauge-field production can lead to a considerable enhancement in the amplitude of the scalar power spectrum, consistent with prior observations of this kind [109]. Such an enhancement can be linked to important phenomenological implications, for instance an increased rate of primordial-black-hole formation after inflation [253], or the sourcing of scalar-induced gravitational waves (SIGWs) at second order in perturbation theory [248, 249].

From this perspective, the results we have presented in this chapter are an important technical milestone, as both of these phenomenological effects are likely most relevant in scenarios where gauge-field production in axion inflation is efficient enough to also effectively disrupt the slow-roll attractor. In particular, the self-consistent method for computing the scalar power spectrum based on the background dynamics obtained with the GEF can be used in the future to study both PBH production and SIGW production for axion inflation in the strong-backreaction regime, albeit while neglecting to account for gauge-field-backreaction effects on scalar perturbations as long as their effect is still ill-understood. Moreover, the numerical strategy we have devised here also entails a secondary benefit for the purposes of this thesis, as it can also be readily applied to the study of gauge-field-induced tensor perturbations in axion inflation, as we shall do in the next chapter.

Chapter 9

Gravitational waves from axion inflation

This chapter is based on Refs. [4, 5]. Detailed contributions of the author to these publications are outlined at the beginning of this thesis.

In the final chapter of this thesis, we turn our attention to the study of gravitational waves (GW) from axion inflation with a focus on gravitational radiation sourced by gauge fields. This is qualitatively similar to the sourcing of scalar perturbations from gauge fields discussed in the previous chapter, with the difference being that we are here concerned with the production of tensor perturbations.

Gauge-field-induced tensor perturbations contribute to the inflationary background of GWs from axion inflation and can enhance its amplitude above that of inflationary vacuum GWs [56–58]. This enhancement of the inflationary GW spectrum is an important prediction of axion inflation which sets the model apart from the conventional slow-roll model. As explained in chapter 1, single-field slow-roll inflation predicts a nearly scale-invariant spectrum of inflationary vacuum GWs, which, in combination with current CMB limits on the amplitude of inflationary GWs, entails that this GW background is likely not detectable by any current or next-generation GW experiments. In contrast, gauge-field-induced gravitational waves (GFIGWs) that are created during axion inflation can amplify the inflationary gravitational-wave background such that it could be observed with current or next-generation GW observatories [51, 82–84].

Several analytical studies of axion inflation suggest that an observable stochastic gravitational-wave background (SGWB) from axion inflation requires that the inflaton field sources a strongly enhanced background of electromagnetic radiation that can then amply produce GWs [51, 82, 84]. However, as we are well familiar with from previous chapters, efficient gauge-field production can also entail strong gauge-field backreaction, which can drastically alter the overall dynamics of axion inflation, and cannot be accounted for using analytical means. Recently, several authors have therefore performed numerical analyses of GFIGW production during axion inflation to also capture strong-backreaction effects. For instance, Refs. [83, 87, 88] computed GW spectra by relying on the mode-by-mode approach to resolve the dynamics of axion inflation, while Ref. [380] used lattice simulations to study the efficiency of GFIGW production in this model. Other authors also used lattice techniques to study GFIGW production from a phase of preheating after axion inflation [80, 385, 386]. These works found that GFIGW production during axion inflation can also be too efficient, resulting in an overproduction of GWs in conflict with CMB and BBN bounds on additional radiation beyond SM photons and neutrinos [380, 385, 386]. In this chapter, we highlight our contribution to these ongoing numerical efforts, relying on the gradient-expansion formalism (GEF) as a basis for computing the spectrum of GWs from axion inflation.

In our analysis, we not only account for potential strong-backreaction effects, but we also take into consideration how Schwinger pair creation affects the dynamics of axion inflation. As we have seen in chapter 6, the strong electromagnetic fields in this model can efficiently create charge carriers from the vacuum, which then also drain energy from the electromagnetic background. This in turn reduces the ability for electromagnetic fields to further source GWs, an effect which has thus far only been estimated using analytical means [121]. Relying on our improved effective

modelling of Schwinger pair creation from chapter 6, we now expand upon these estimates using the GEF. To our knowledge, this presents the first fully numerical study of GFIGW production that also accounts for the Schwinger effect.

In this chapter, we therefore study GW production in two variants of axion inflation. The first is the conventional model of axion inflation, in which Schwinger pair creation is not taken into account, a model we refer to as *pure axion inflation* (PAI for short). The second model we consider is *fermionic axion inflation* (FAI), which accounts for Schwinger pair production of fermion during axion inflation. In particular, we work with the variant of FAI in which the inflaton field couples to the Standard Model (SM) hypercharge field, and the Schwinger effect leads to the production of SM fermions. As we have seen in chapter 6, Schwinger pair creation can be highly efficient in this model, and it can thus have a pronounced impact on GFIGW production.

For both these models, we assess the detectability of GFIGWs in next-generation observatories like LISA or ET, and for current PTA experiments like NANOGrav. To do so, we model the dynamics of inflation by assuming a quadratic potential for the inflaton field, which represents a realistic possibility for the late-time inflaton potential away from CMB scales. Therefore, our model is described by two parameters, the inflaton mass, m , describing the quadratic potential, and the axion–vector coupling β , which regulates the efficiency of gauge-field production and backreaction during axion inflation. Based on this modelling, we perform a comprehensive parameter scan for both variants of axion inflation, PAI and FAI, to determine regions of parameter scan where gauge-field production can lead to an observable signal in GFIGWs. To do so, we also consistently take into account both cosmological constraints and upper bounds on the amplitude of GWs suggested by the non-observation of an SGWB by LIGO and Virgo [239].

For the case of pure axion inflation, our study is unique in that other authors have only studied GFIGW production for a particular set of benchmark points, while the numerical efficiency of the GEF allows us to perform a more thorough exploration of our model parameter space. In contrast, our analysis of GW production for fermionic axion inflation is the first of its kind, and our GEF-based approach permits us to perform an elaborate analysis concerning the dynamical impact of fermions on the sourcing of GWs in this model.

Our discussion in this chapter is structured as follows. In section 9.1, we present the derivation of the GW spectrum for our two models of axion inflation. Following this, section 9.2 details our modelling assumptions for PAI and FAI, as well as introducing experimental constraints for our analysis, and the method by which we characterize the detectability of GWs. We then proceed with the discussion of our results. First, we consider PAI in section 9.3, followed by FAI in section 9.4. Finally, we summarize our findings in section 9.5, and give an outlook towards future research directions motivated by our results.

9.1 Gauge-field-induced gravitational waves

We begin by establishing how we compute the spectrum of inflationary GWs for our two variants of axion inflation, PAI and FAI. To this end, we first discuss the possible sources for GWs in these two models, and motivate why we focus on gauge fields as the main source of interest. We then present the derivation of the tensor power spectrum for these two models by accounting for both vacuum and gauge-field-induced tensor perturbations. This derivation yields a general expression for the inflationary tensor power spectrum which we can use to numerically determine the GW spectrum at the end of inflation for our two models of axion Inflation. We then review how one can account for the evolution of GWs through cosmic history to relate the tensor power spectrum at the end of inflation to the present-day spectrum of inflationary GWs.

9.1.1 Tensor perturbations in axion inflation

Our starting point is a generic action for axion inflation which we can use to both describe the model of pure axion inflation (PAI) and fermionic axion inflation (FAI):

$$S = \int d^4x \sqrt{-g} \left[-\frac{M_{\text{P}}^2}{2} R + \frac{1}{2} g^{\mu\nu} \partial_\mu \phi \partial_\nu \phi - V(\phi) - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} - \frac{\beta}{4M_{\text{P}}} \phi F_{\mu\nu} \tilde{F}^{\mu\nu} + \mathcal{L}_\chi \right]. \quad (9.1)$$

The symbols appearing therein have their conventional interpretation within the confines of this thesis: R is the Ricci scalar, $g_{\mu\nu}$ is the metric, and g is its determinant; the inflaton field is given by ϕ with $V(\phi)$ its potential; $F_{\mu\nu}$ is the field-strength tensor to the gauge field A_μ , and $\tilde{F}_{\mu\nu}$ is the dual to the field-strength tensor; β denotes the dimensionless axion–vector coupling. Following our description in chapter 6, we have also allowed for an interaction between the gauge field A_μ and a charged sector which is represented by the unspecified Lagrangian $\mathcal{L}_\chi$. Pure axion inflation simply entails that $\mathcal{L}_\chi = 0$.

Relevant for our purposes is the energy–momentum tensor that one obtains from this action, which is (cf. chapter 6)

$$T_{\mu\nu} = \partial_\mu \phi \partial_\nu \phi - g^{\alpha\beta} F_{\mu\alpha} F_{\nu\beta} - g_{\mu\nu} \left[\frac{1}{2} \partial_\alpha \phi \partial^\alpha \phi - V(\phi) - \frac{1}{4} F_{\alpha\beta} F^{\alpha\beta} \right] + (T_\chi)_{\mu\nu}, \quad (9.2)$$

where $(T_\chi)_{\mu\nu}$ denotes the possible contribution of charge carriers to $T_{\mu\nu}$. In particular, we later focus on the case where these charge carriers are SM fermions.

Our focus is on the production and evolution of tensor perturbations in axion inflation, thus we make the Ansatz of a perturbed FLRW metric for $g_{\mu\nu}$,

$$g_{\mu\nu} dx^\mu dx^\nu = a^2(\eta) \left[d\eta^2 - (\delta_{ij} + h_{ij}^{\text{TT}}(\eta, \mathbf{x})) dx^i dx^j \right], \quad (9.3)$$

with conformal time η and the scale factor a , as usual. Since we want to study GWs, we only keep the relevant transverse–traceless metric perturbations h_{ij}^{TT} ; we shall work to linear order in perturbation theory, so we can disregard scalar and vector perturbations in what follows.

For the homogeneous and isotropic background, we follow the discussion of axion inflation outlined in chapters 3 and 6. Concerning the inflaton field, we assume the background FLRW spacetime is only affected by the inflaton’s homogeneous and isotropic vacuum expectation value, $\varphi(\eta) = \langle \phi(\eta, \mathbf{x}) \rangle$, treating deviations from φ as perturbations, $\delta\phi(\eta, \mathbf{x}) = \phi(\eta, \mathbf{x}) - \varphi(\eta)$. The gauge field is described in terms of electric- and magnetic three-fields, $\mathbf{E}$ and $\mathbf{B}$, whose components we define as $F_{0i} = a^2 E^i$ and $F_{ij} = -a^2 \varepsilon_{ijk} B^k$ (cf. Eq. (3.10b)). These fields can backreact on the inflationary background through the zero-modes $\langle \mathbf{E}^2 \rangle$, $\langle \mathbf{B}^2 \rangle$, and $\langle \mathbf{E} \cdot \mathbf{B} \rangle$. Finally, if charge carriers are present in the system, $\mathcal{L}_\chi \neq 0$, they are treated effectively as outlined in chapter 6. Moreover, in what follows, we are implicitly assuming a phase of inflationary expansion driven by the inflaton’s potential energy density, $V(\varphi)$. In particular, this modelling allows us to treat the entire dynamics of the inflationary background using the GEF as outlined in chapter 4 for PAI ($\mathcal{L}_\chi = 0$), and chapter 6 for FAI ($\mathcal{L}_\chi \neq 0$).

Our attention in this section shall then be on the tensor perturbations h_{ij}^{TT} in Eq. (9.3). From the linearized Einstein equations, it follows that h_{ij}^{TT} is sourced by the anisotropic stress, Σ_{ij} , of the energy–momentum tensor in Eq. (9.2) (cf. Eq. (7.31)):

$$(h_{ij}^{\text{TT}})'' + 2\mathcal{H} (h_{ij}^{\text{TT}})' - \Delta h_{ij}^{\text{TT}} = \frac{2}{M_{\text{P}}^2} a^2 (\Sigma_{ij})^{\text{TT}}. \quad (9.4)$$

Here, $\mathcal{H} = a'/a$ is the comoving Hubble rate, and the subscript “TT” on the right-hand side denotes that only the transverse-traceless part of Σ_{ij} may source tensor perturbations.

For PAI and FAI, the anisotropic stress can be computed from Eq. (9.2), and is given by three constituents:

$$\Sigma_{ij} = (\Sigma_\phi)_{ij} + (\Sigma_{\text{EM}})_{ij} + (\Sigma_\chi)_{ij} . \quad (9.5)$$

The first piece comes from the inflaton field,

$$(\Sigma_\phi)_{ij} = \partial_i \phi \partial_j \phi , \quad (9.6)$$

the second term constitutes the anisotropic stress due to gauge fields,

$$(\Sigma_{\text{EM}})_{ij} = -(E_i E_j + B_i B_j) , \quad (9.7)$$

and the final piece, $(\Sigma_\chi)_{ij}$, is then the anisotropic stress due to charge carriers, which is only relevant for the model of FAI (PAI assumes charge carriers are absent).

The inflaton’s anisotropic stress, $(\Sigma_\phi)_{ij}$, vanishes at linear order in perturbation theory as long as $\delta\phi$ remains perturbatively small.¹ It should however be noted that, in the case of strong gauge-field backreaction, inflaton inhomogeneities can be abundantly sourced by the gauge field, as we have discussed in chapter 4. In this case, the inflation’s anisotropic stress could become a relevant source of GWs. However, when this happens, also our perturbative Ansatz for the matter sector breaks down, as inflaton inhomogeneities start to meaningfully affect the overall dynamics of the background FLRW spacetime. Thus, to be consistent with our perturbative approach, we do not account for the anisotropic stress due to inflaton inhomogeneities in our subsequent analysis.

We also do not explicitly compute the production of tensor perturbations due to the anisotropic stress from charge carriers, $(\Sigma_\chi)_{ij}$. Although we do account for the creation of charge carriers via Schwinger pair creation and their subsequent backreaction on the background dynamics, we do so by means of the effective-model approach presented in chapter 6. In this approach, we cannot infer the degree of anisotropy in the charged sector, so we cannot directly compute how efficiently $(\Sigma_\chi)_{ij}$ sources tensor perturbations. We shall however estimate the relevance of this contribution in later parts of this chapter.

After these considerations, all we are therefore left with is the electric and magnetic fields acting as a source of tensor perturbations. Equation (9.4) then reads

$$(h_{ij}^{\text{TT}})'' + 2\mathcal{H} (h_{ij}^{\text{TT}})' - \Delta h_{ij}^{\text{TT}} = -\frac{2}{M_{\text{P}}^2} a^2 (E_i E_j + B_i B_j)^{\text{TT}} . \quad (9.8)$$

Starting from this equation, we now proceed to derive the conventional tensor power spectrum for axion inflation following the discussions in Refs. [56–58, 83].

9.1.2 Calculation of the tensor power spectrum

The derivation of the tensor power spectrum is similar in spirit to the one for the scalar power spectrum given in the previous chapter. Similar to this calculation, we therefore first go from the position-space description of h_{ij}^{TT} to Fourier space by writing

$$h_{ij}^{\text{TT}}(\eta, \mathbf{x}) = \int \frac{d^3 \mathbf{k}}{(2\pi)^{3/2}} \sum_{\lambda=\pm 1} \left(\epsilon_i^\lambda(\mathbf{k}) \epsilon_j^\lambda(\mathbf{k}) h_{\mathbf{k},\lambda}(\eta) e^{i\mathbf{k}\cdot\mathbf{x}} + \text{h.c.} \right) . \quad (9.9)$$

¹At second order in perturbation theory, inflaton perturbations contribute to the sourcing of scalar-induced gravitational waves (cf. chapter 1 and Ref. [249] for a review thereof).

Here, we have chosen to represent the two propagating degrees of freedom in h_{ij}^{TT} in a circular-polarization basis with the polarization vectors $\epsilon^\lambda(\mathbf{k})$ defined via the usual relationships

$$\mathbf{k} \cdot \epsilon^\lambda(\mathbf{k}) = 0, \quad [i\mathbf{k} \times \epsilon^\lambda(\mathbf{k})] = \lambda k \epsilon^\lambda(\mathbf{k}), \quad \epsilon^{\lambda*}(\mathbf{k}) \cdot \epsilon^{\lambda'}(\mathbf{k}) = \delta^{\lambda\lambda'}. \quad (9.10)$$

In our case, the circular-polarization basis is a convenient choice for decomposing h_{ij}^{TT} , since axion inflation sources helical gauge fields, which therefore source helical tensor perturbations via the source term in Eq. (9.8). The Fourier amplitudes $h_{\mathbf{k},\lambda}$ then evolve according to

$$h''_{\mathbf{k},\lambda} + 2\mathcal{H}h'_{\mathbf{k},\lambda} + k^2 h_{\mathbf{k},\lambda} = -\frac{2}{M_{\text{P}}^2} a^2 \Pi_{\lambda}^{ij}(\mathbf{k}) \int \frac{d^3\mathbf{x}}{(2\pi)^{3/2}} (E_i E_j + B_i B_j) e^{-i\mathbf{k}\cdot\mathbf{x}}, \quad (9.11)$$

where $\Pi_{\lambda}^{ij} = \epsilon_{-\lambda}^i(\mathbf{k}) \epsilon_{-\lambda}^j(\mathbf{k})$ is the projector onto the circular polarization basis for $\lambda = \pm 1$. We also used that projecting onto the circular-polarization basis is equivalent to projecting onto the transverse–traceless component of a tensor (see, e.g., Ref. [58]), thus we can drop the superscript TT for the source term on the right-hand side of Eq. (9.11).

A solution to this equation can now be obtained by combining its homogeneous and particular solutions,²

$$h_{\mathbf{k},\lambda}(\eta) = \frac{2}{M_{\text{P}}} \left(Q_{\mathbf{k},\lambda}^{\text{vac}}(\eta) + Q_{\mathbf{k},\lambda}^{\text{ind}}(\eta) \right). \quad (9.12)$$

Here, $Q_{\mathbf{k},\lambda}^{\text{vac}}$ corresponds to the homogeneous solution, while $Q_{\mathbf{k},\lambda}^{\text{ind}}$ gives the particular solution to Eq. (9.11). Analogous to the case of curvature perturbations in the previous chapter, $Q_{\mathbf{k},\lambda}^{\text{vac}}$ encodes the conventional vacuum tensor perturbations that are also present in the case of standard slow-roll inflation, while $Q_{\mathbf{k},\lambda}^{\text{ind}}$ can be interpreted as tensor perturbations induced by electromagnetic fields.

The discussion of the vacuum contribution proceeds similar to the one of vacuum scalar perturbations. We treat $Q_{\mathbf{k},\lambda}^{\text{vac}}$ as a quantum field which we can decompose into a mode function, $u_{\lambda}(\eta, k)$, and a set of creation and annihilation operators, $\hat{b}_{\mathbf{k},\lambda}^{\dagger}$ and $\hat{b}_{\mathbf{k},\lambda}$,

$$Q_{\mathbf{k},\lambda}^{\text{vac}} = u_{\lambda}(\eta, k) \hat{b}_{\mathbf{k},\lambda} + u_{\lambda}^*(\eta, k) \hat{b}_{-\mathbf{k},\lambda}^{\dagger}. \quad (9.13)$$

The ladder operators obey canonical commutation relations

$$[\hat{b}_{\mathbf{k},\lambda}, \hat{b}_{\mathbf{k}',\lambda'}^{\dagger}] = \delta_{\lambda,\lambda'} \delta^{(3)}(\mathbf{k} - \mathbf{k}'), \quad [\hat{b}_{\mathbf{k},\lambda}, \hat{b}_{\mathbf{k}',\lambda'}] = [\hat{b}_{\mathbf{k},\lambda}^{\dagger}, \hat{b}_{\mathbf{k}',\lambda'}^{\dagger}] = 0, \quad (9.14)$$

and the mode functions solve

$$u_{\lambda}'' + 2\mathcal{H}u_{\lambda}' + k^2 u_{\lambda} = 0. \quad (9.15)$$

For modes deep inside the horizon, Bunch–Davies initial conditions are then realized by

$$u_{\lambda}(\eta, k) \simeq \frac{1}{a(\eta)} \frac{1}{\sqrt{2k}} e^{-ik\eta}, \quad \text{for } k \gg \mathcal{H}, \quad (9.16)$$

which uniquely fixes the vacuum state annihilated by $b_{\mathbf{k},\lambda}$ to be the Bunch–Davies vacuum.

²The factor $2/M_{\text{P}}$ in Eq. (9.12) needs to be included such that the second-order perturbed Einstein–Hilbert action expressed through the Fourier amplitude $^{M_{\text{P}}/2}h_{\mathbf{k},\lambda}$ contains a canonically normalized kinetic term, see e.g., the textbook [192].

Next, let us turn to the induced contribution. It may again be determined by means of the Green function method and is given by

$$Q_{\mathbf{k},\lambda}^{\text{ind}}(\eta) = -\frac{1}{M_{\text{P}}} \int_{-\infty}^0 d\tau a^2(\tau) G_k(\eta, \tau) \Pi_{\lambda}^{ij}(\mathbf{k}) \int \frac{d^3\mathbf{x}}{(2\pi)^{3/2}} (E_i E_j + B_i B_j) e^{-i\mathbf{k}\cdot\mathbf{x}}. \quad (9.17)$$

Here, $G_k(\eta, \tau)$ is the retarded Green function which solves

$$(\partial_{\eta}^2 + 2\mathcal{H}\partial_{\eta} + k^2) G_k(\eta, \tau) = \delta(\eta - \tau), \quad (9.18)$$

for the boundary conditions

$$G_k(\eta, \tau \geq \eta) \equiv 0, \quad \left. \frac{\partial}{\partial \eta} G_k(\eta, \tau) \right|_{\tau \rightarrow \eta} = 1. \quad (9.19)$$

The retarded Green function can again be constructed using two linearly independent solutions to Eq. (9.15), for instance the vacuum mode function $u_{\lambda}(\eta, k)$ of $Q_{\mathbf{k},\lambda}^{\text{vac}}$ in Eq. (9.13) and its complex conjugate. In this case, one obtains

$$G_k(\eta, \tau) = \frac{\text{Im}[u_{\lambda}^*(\eta, k) u_{\lambda}(\tau, k)]}{\text{Im}[u_{\lambda}(\tau, k) u_{\lambda}^*(\tau, k)]} \theta(\eta - \tau). \quad (9.20)$$

As in the previous chapter, the gauge-field operators E_i and B_i appearing as a source in Eq. (9.17) may be expressed in terms of the gauge-field mode functions $A_{\lambda}(\eta, k)$, circular polarization vectors $\epsilon^{\lambda}(\mathbf{k})$, and the associated ladder operators for the gauge field, $\hat{a}_{\mathbf{k},\lambda}^{\dagger}$ and $\hat{a}_{\mathbf{k},\lambda}$ (cf., e.g., Eqs. (8.32) and (8.33)). These creation/annihilation operators commute with the ones for the vacuum tensor modes $\hat{b}_{\mathbf{k},\lambda}^{\dagger}$ and $\hat{b}_{\mathbf{k},\lambda}$ in Eq. (9.13). The induced Fourier amplitudes $Q_{\mathbf{k},\lambda}^{\text{ind}}$ are then quadratic in the gauge-field ladder operators, $\hat{a}_{\mathbf{k},\lambda}$ and $\hat{a}_{\mathbf{k},\lambda}^{\dagger}$, and commute with the vacuum Fourier amplitudes $Q_{\mathbf{k},\lambda}^{\text{vac}}$.

Continuing from here, one can then define a tensor power spectrum for each circular polarization of the Fourier amplitudes $h_{\mathbf{k},\lambda}$;

$$\langle h_{\mathbf{k},\lambda}(\eta) h_{\mathbf{k}',\lambda'}(\eta) \rangle = \delta^{(3)}(\mathbf{k} + \mathbf{k}') \delta_{\lambda\lambda'} \frac{2\pi^2}{k^3} \mathcal{P}_{h,\lambda}(\eta, k). \quad (9.21)$$

The total tensor power spectrum defined in chapter 1 is simply the sum over the two circular polarizations,

$$\mathcal{P}_h(\eta, k) = \sum_{\lambda=\pm 1} \mathcal{P}_{h,\lambda}(\eta, k). \quad (9.22)$$

Expanding $h_{\mathbf{k},\lambda}$ into a vacuum and an induced contribution as in Eq. (9.12), we can apply Wick's theorem to find that the power spectrum $\mathcal{P}_{h,\lambda}$ separates into two contributions

$$\mathcal{P}_{h,\lambda}(\eta, k) = \mathcal{P}_{h,\lambda}^{\text{vac}}(\eta, k) + \mathcal{P}_{h,\lambda}^{\text{ind}}(\eta, k). \quad (9.23)$$

The vacuum power spectrum, $\mathcal{P}_{h,\lambda}^{\text{vac}}(\eta, k)$, follows from $Q_{\mathbf{k},\lambda}^{\text{vac}}$ and is given by

$$\mathcal{P}_{h,\lambda}^{\text{vac}}(\eta, k) = \frac{2k^3}{\pi^2 M_{\text{P}}^2} |u_{\lambda}(\eta, k)|^2. \quad (9.24)$$

Vacuum GW modes are unpolarized, since the mode equation for the modes u_{λ} in Eq. (9.15) is independent of polarization, as are the boundary conditions in Eq. (9.16). Therefore, the vacuum

power spectrum is also unpolarized, $\mathcal{P}_{h,+}^{\text{vac}} = \mathcal{P}_{h,-}^{\text{vac}}$. The induced power spectrum, $\mathcal{P}_{h,\lambda}^{\text{ind}}$, is computed from $Q_{k,\lambda}^{\text{ind}}$. After applying Wick's theorem, one obtains

$$\mathcal{P}_{T,\lambda}^{\text{ind}}(\eta, k) = \frac{k^3}{4\pi^2 M_{\text{P}}^4} \int \frac{d^3 \mathbf{p}}{(2\pi)^3} \sum_{\alpha, \beta = \pm 1} \left(1 + \lambda \alpha \frac{\mathbf{k} \cdot \mathbf{p}}{kp}\right)^2 \left(1 + \lambda \beta \frac{k^2 - \mathbf{k} \cdot \mathbf{p}}{kq}\right)^2 \quad (9.25)$$

$$\times \left| \int_{-\infty}^0 d\tau \frac{G_k(\eta, \tau)}{a^2(\tau)} [A'_\alpha(\tau, p) A'_\beta(\tau, q) + \alpha \beta pq A_\alpha(\tau, p) A_\alpha(\tau, q)] \right|^2,$$

where $q = |\mathbf{k} - \mathbf{p}|$, $k = |\mathbf{k}|$, and $p = |\mathbf{p}|$.³ In the case of PAI, the gauge-field mode functions $A_\lambda(\eta, k)$ solve the standard mode equation for axion inflation, Eq. (3.18), while in the case of FAI, they solve the mode equation which accounts for the presence of a conducting medium of charge carriers, Eq. (6.35). In either case, one polarization mode is enhanced over the other one, and thus, the induced power spectrum $\mathcal{P}_{h,\lambda}^{\text{ind}}$ is polarized; $\mathcal{P}_{h,+}^{\text{ind}} \neq \mathcal{P}_{h,-}^{\text{ind}}$.

Assuming a period of slow-roll inflation and neglecting slow-roll corrections, one may analytically determine the mode functions $u_\lambda(\eta, k)$ and the Green function $G_k(\eta, \tau)$ similar to the case of scalar perturbations given in the previous chapter. This yields the standard result for the tensor power spectrum in the super-horizon limit (cf. Eq. (1.82)):

$$\mathcal{P}_h^{\text{vac}} = \mathcal{P}_{h,+}^{\text{vac}} + \mathcal{P}_{h,-}^{\text{vac}} \simeq 2 \left(\frac{H}{\pi M_{\text{P}}} \right)^2, \quad (9.26)$$

Furthermore, for PAI in the weak-backreaction regime, the induced tensor power spectrum can be computed analytically using the approximate gauge-mode functions $A_\lambda(\eta, k)$ in Eq. (3.30) for the case of a constant instability parameter $\xi = (\beta \dot{\phi})/(2H M_{\text{P}})$. Assuming, for definitiveness, that $\xi > 0$, this analytical computation gives [56–58]

$$\mathcal{P}_{h,\lambda}^{\text{ind}} \simeq c_\lambda \left(\frac{H}{M_{\text{P}}} \right)^4 \frac{\exp(4\pi\xi)}{\pi^2 \xi^6}, \quad (9.27)$$

with the coefficients being $c_+ \simeq 8.6 \times 10^{-7}$ and $c_- \simeq 1.8 \times 10^{-9}$ (see Ref. [58]). In the case of FAI, no such analytical results for $\mathcal{P}_{h,\lambda}^{\text{ind}}$ are known.

For our purposes, these analytical results however merely serve the purpose of illustration. In what follows, we instead compute both the vacuum and induced tensor power spectrum numerically. The strategy to do so is analogous to the one we used to compute the induced scalar power spectrum in Sec. 8.3. In summary, it entails; a) computing the full dynamical background evolution using the GEF method; b) using the GEF solution to determine the mode functions $A_\lambda(\eta, k)$, the tensor mode functions $u_\lambda(\eta, k)$, and the Green function $G_k(\eta, \tau)$ numerically; c) using these to determine both $\mathcal{P}_{h,\lambda}^{\text{vac}}$ from Eq. (9.24) and $\mathcal{P}_{h,\lambda}^{\text{ind}}$ by performing the integration in Eq. (9.25) numerically. For details on this procedure, we refer to App. B.

9.1.3 From inflationary gravitational waves until today

The calculation we have performed above only serves to determine the spectrum of tensor perturbations that is created during inflation. However, what we are interested in for phenomenological applications is how these inflationary tensor perturbations translate to the spectrum of GWs which

³The momentum integral over the gauge-field mode functions in Eq. (9.25) diverges for $p \rightarrow \infty$ just as the induced scalar power spectrum in the previous chapter. As usual, we regularize this ultraviolet (UV) divergence by means of the conventional UV cutoff at the tachyonic instability scale k_{h} [83]. Depending on the model in questions, this scale is either given by Eq. (3.24) for PAI or Eq. (6.56) for FAI.

could be detected by current or future experiments today. Since axion inflation produces tensor perturbations from quantum fields, such GWs would then be detectable today as a stochastic gravitational-wave background (SGWB) (see the textbook [192] or the review article [219] for more details on SGWBs from inflation and other cosmological sources).

The spectrum of a cosmological SGWB is conventionally expressed in terms of the GW energy density in units of the critical energy density and per logarithmic frequency,

$$\Omega_{\text{GW}}(f) = \frac{1}{3H_0^2 M_{\text{P}}^2} \frac{d\rho_{\text{GW}}(f)}{d \ln f}, \quad (9.28)$$

where $H_0 = h \times 100 \text{ km s}^{-1} \text{ Mpc}^{-1}$ is the Hubble rate today, f is the frequency of GWs today, and ρ_{GW} is the present-day energy density in GWs, i.e., [192],

$$\rho_{\text{GW}} = \frac{1}{4} M_{\text{P}}^2 \langle \dot{h}_{ij}^{\text{TT}}(\eta_0, \mathbf{x}) \dot{h}^{ij, \text{TT}}(\eta_0, \mathbf{x}) \rangle, \quad (9.29)$$

with η_0 denoting conformal time today. $\Omega_{\text{GW}}(f)$ can then be directly related to the tensor power spectrum we have introduced in the previous section by accounting for two effects: (i) the evolution of a GW mode through FLRW spacetime after inflation has ended, and the GW source has switched off (in our case, this means the electromagnetic fields have decayed), and (ii) the cosmological redshift of the GW frequency (see, e.g., Ref. [192]).

For the first effect, one considers that, following its creation during inflation, a GW mode $h_{\mathbf{k},\lambda}$ with wavenumber $k = |\mathbf{k}|$ exits the Hubble horizon at a time $\eta_{\text{out},k}$ defined by $k \simeq \mathcal{H}(\eta_{\text{out},k})$. While this mode is super-horizon, we can assume that the corresponding tensor power spectrum is conserved at its value upon horizon exit, $\mathcal{P}_h(\eta_{\text{out},k}, k)$. As usual, once inflation has ended, $\mathcal{H}$ begins to decrease, and eventually the GW mode with wavenumber k reenters the horizon at a time $\eta_{\text{in},k}$ with $k \simeq \mathcal{H}(\eta_{\text{in},k})$. Since the tensor power spectrum is assumed to be conserved while k is super-horizon, we have that $\mathcal{P}_h(\eta_{\text{out},k}, k) \simeq \mathcal{P}_h(\eta_{\text{in},k}, k)$. After the horizon reentry of $h_{\mathbf{k},\lambda}$, we assume that the GW source (the gauge field) has become inactive, and the subsequent evolution of $h_{\mathbf{k},\lambda}$ is governed by the source-free evolution equation

$$h_{\mathbf{k},\lambda}'' + 2\mathcal{H}h_{\mathbf{k},\lambda}' + k^2 h_{\mathbf{k},\lambda} = 0. \quad (9.30)$$

One can then relate the GW mode at horizon reentry, $h_{\mathbf{k},\lambda}(\eta_{\text{in},k})$, to its value today, $h_{\mathbf{k},\lambda}(\eta_0)$, by solving this equation for the known cosmological evolution through radiation domination and matter domination. The effect can be captured by means of a transfer function relating the two amplitudes; $h_{\mathbf{k},\lambda}(\eta_0) = \mathcal{T}_{\text{GW}}(k) h_{\mathbf{k},\lambda}(\eta_{\text{in},k})$. From this, it immediately follows that

$$\mathcal{P}_h(\eta_0, k) = |\mathcal{T}_{\text{GW}}(k)|^2 \mathcal{P}_h(\eta_{\text{in},k}, k) \simeq |\mathcal{T}_{\text{GW}}(k)|^2 \mathcal{P}_h(\eta_{\text{out},k}, k). \quad (9.31)$$

One can then switch from comoving wavenumber k to the physical frequency f measured today by the standard relationship $f = k/(2\pi a_0)$ with a_0 being the scale factor today. This accounts for the cosmological frequency redshift. Putting everything together, one then obtains the following relationship between Ω_{GW} in Eq. (9.28) and $\mathcal{P}_h(\eta_{\text{out},k}, k)$ [192];

$$\Omega_{\text{GW}}(f) = \frac{\pi^2}{3H_0^2} f^2 |\mathcal{T}_{\text{GW}}(f)|^2 \mathcal{P}_h(\eta_{\text{out},k_f}, k_f), \quad k_f = 2\pi a_0 f. \quad (9.32)$$

To obtain this result, one uses that $d\rho_{\text{GW}}/d \log f \propto f^2 \mathcal{P}_h(\eta_0, k_f)$ which follows from spatial isotropy and the fact that $k \gg \mathcal{H}$ once $h_{\mathbf{k},\lambda}$ has reentered the horizon (see Ref. [192]).

In this chapter, we use the following standard expression for the transfer function [219];

$$|\mathcal{T}_{\text{GW}}(f)|^2 \simeq \frac{H_0^2 \Omega_r}{8\pi^2 f^2} \frac{g_*(T_f)}{g_*(T_0)} \left(\frac{g_{*,S}(T_0)}{g_{*,S}(T_f)} \right)^{4/3} \left[1 + \frac{9}{16} \left(\frac{f_{\text{eq}}}{\sqrt{2}f} \right)^2 \right] |\mathcal{T}_{\text{reh}}(f)|^2. \quad (9.33)$$

Here, the functions $g_*(T)$ and $g_{*,S}(T)$ represent the effective number of relativistic degrees of freedom which, respectively, contribute to the energy and entropy density of the SM thermal plasma at temperature T . The temperature T_f corresponds to the time when the wavenumber k_f reentered the horizon and $T_0 \simeq 2.73$ K is the temperature of the CMB photons today. As usual, Ω_r is the fractional energy density in radiation today, $h^2 \Omega_r = 4.2 \times 10^{-5}$ [16]. The fourth factor describes the transition of GW modes from radiation to matter domination, with $f_{\text{eq}} = 2.1 \times 10^{-17}$ Hz the frequency corresponding to matter–radiation equality [219]. The last factor, $|\mathcal{T}_{\text{reh}}(f)|^2$, affects modes which reenter the horizon during a period of reheating before the onset of radiation domination. This term is evidently subject to the details of reheating. In our subsequent analysis, we shall assume instantaneous reheating, which corresponds to $|\mathcal{T}_{\text{reh}}(f)|^2 = 1$.⁴

One can also simply expand the relationship $k_f = 2\pi a_0 f$, between a frequency f and the corresponding comoving momentum k_f , following the same logic that lead to Eq. (1.85). Assuming instantaneous reheating, one obtains

$$f = \frac{k_f}{2\pi a_0} = \frac{k_f}{2\pi a_{\text{end}}} \left(\frac{g_{*,S}(T_0)}{g_{*,S}(T_{\text{reh}})} \right)^{1/3} \frac{T_0}{T_{\text{reh}}}, \quad (9.34)$$

where a_{end} denotes the scale factor at the end of inflation, and the reheating temperature T_{reh} , is given by Eq. (1.48). From this, one may also estimate the maximal frequency reached by GWs produced during inflation, which correspond to modes leaving the Hubble horizon at the end of inflation, $k = a_{\text{end}} H_{\text{end}}$,

$$f_{\text{end}} = \frac{H_{\text{end}}}{2\pi} \left(\frac{a_{\text{end}}}{a_{\text{reh}}} \right) \left(\frac{g_{*,S}(T_0)}{g_{*,S}(T_{\text{reh}})} \right)^{1/3} \frac{T_0}{T_{\text{reh}}} \stackrel{\text{inst. reh.}}{\simeq} 1.83 \times 10^{11} \left(\frac{H_{\text{end}}}{M_{\text{P}}} \right)^{1/2} \text{ Hz}. \quad (9.35)$$

Here, we used that $g_*(T_{\text{reh}}) \simeq g_{*,S}(T_{\text{reh}}) \simeq 104.4$ for $T_{\text{reh}} \gtrsim 2 \times 10^2$ GeV [453, 454].

In practice, we can then use Eq. (9.32) to determine Ω_{GW} by relying on numerical results for the tensor power spectrum based on Eqs. (9.23) to (9.25). Furthermore, we can deduce the scale factor at the end of inflation, a_{end} ⁵ from the GEF results for the inflationary background, which we combine with the numerical values for $g_*(T)$ and $g_{*,S}(T)$ from Refs. [453, 454] to evaluate Eqs. (9.33) and (9.34). This workflow allows one to go from the inflationary background dynamics via the tensor power spectrum to Ω_{GW} , and is implemented in the Python package that we developed alongside this thesis; the GEFF [7].

9.2 Model specifications and constraints

Having established how we can compute GW spectra for axion inflation, we can proceed with the phenomenological analysis concerning their detectability with current or next-generation GW experiments. In this section, we first consider technical aspects for this analysis. Specifically, we first specify our treatment of background dynamics for both PAI and FAI, and we consider what constraints CMB measurements impose upon them. Then, we discuss what GW experiments are

⁴See also Refs. [451, 452] for an approximate expression for $\mathcal{T}_{\text{reh}}(f)$ assuming a period of reheating described by an equation of state parameter $w = 0$.

⁵Our convention is to set the scale factor to one when initializing of our system during inflation, $a_{\text{ini}} = 1$.

relevant for our analysis, how we estimate the sensitivity of these experiments to a GW signal, and we discuss existing experimental constraints on the amplitude of a cosmological SGWB.

9.2.1 Modelling of inflationary dynamics and CMB constraints

As discussed in chapter 1, an inflationary background of GWs can extend over many orders of magnitude in frequency, from GWs that have exited the horizon at the time that CMB scales left the inflationary horizon, corresponding to frequencies $f_{\text{CMB}} \sim 10^{-18}$ Hz, to those that have exited the horizon at the end of inflation, with frequencies f_{end} given by Eq. (9.35). In our analysis, we want to capture the entire GW spectrum over all frequencies relevant for GW detection, starting from nanohertz frequencies relevant for PTA experiments. However, we need to also consider that the CMB imposes an upper bound on inflationary GWs at f_{CMB} . To cover the entire frequency range we are ultimately interested in, we want to determine the dynamics of inflation for at least 60 e -foldings.

To cover this entire time-span, we use the GEF to compute the evolution of the inflationary background for both PAI and FAI. From our prior discussions in e.g., chapter 3, we know that the CMB constrains early gauge-field production, as these source non-Gaussian scalar perturbations. Thus, we initialize the GEF using slow-roll initial conditions (cf. chapter 4), i.e., we initialize the inflaton field on the slow-roll attractor of its potential, and we assume gauge-field production is initially absent. Specifically, we model inflation based on a quadratic potential, $V(\phi) = m^2 \phi^2 / 2$, and our initial conditions for the inflaton field are specified accordingly; $\varphi(0) = 15.55 M_{\text{P}}$ and $\dot{\varphi}(0) = -\sqrt{2/3} m M_{\text{P}}$. This way, we are guaranteed 60 e -folds of inflation with the possibility of additional e -foldings because of late-time strong-backreaction effects.

For the case of FAI, we focus on the scenario where the gauge field coupled to the inflaton is identified with the SM hypercharge field. We model this precisely as detailed in Sec. 6.4. This means, we account for Schwinger pair production of all SM fermion species, and we assume that the electroweak symmetry is intact during inflation, such that these fermion fields are massless. Since we assume gauge-field production is inefficient upon initializing our system, we also assume that Schwinger pair production of fermions is initially ineffective, thus, we initialize their energy density as $\rho_{\chi}(0) = 0$. Schwinger pair creation is modelled by means of the mixed-picture representation of the Schwinger-induced current, and we take Schwinger damping as only affecting gauge-field modes with momenta below the instability scale, $k \leq k_{\text{h}}$.

Our choice of potential requires further explanation, given that a quadratic inflation potential is ruled out by CMB measurements (see Sec. 1.5). We nonetheless adhere to this potential for two reasons. First, this potential is the primary benchmark scenario in the literature on strong gauge-field backreaction in axion inflation (see, e.g., Refs. [109, 128, 378]). Second, it yields representative results for GFIGW spectra for any inflationary potential ending its slow-roll trajectory in a nearly quadratic potential. This is due to the large separation between frequencies covered by current and planned GW observatories, $f_{\text{GW}} \sim 10^{-9} - 10^6$ Hz (see below), and the frequency scales constrained by the CMB measurements, $f_{\text{CMB}} \sim 10^{-18}$ Hz. Specifically, we consider that our results are relevant for the class of inflationary models with scalar potentials of approximately the form

$$V(\varphi) = \begin{cases} \frac{1}{2} m^2 \varphi^2 & , \quad |\varphi| \lesssim \varphi_{\text{thr}} \\ V_{\text{CMB}}(\varphi) & , \quad |\varphi| \gtrsim \varphi_{\text{thr}} \end{cases} \quad (9.36)$$

i.e., for models where the potential has an arbitrary shape $V_{\text{CMB}}(\varphi)$ at early times during inflation, such that the potential satisfies CMB constraints, and a quadratic potential as the inflaton approaches the end of inflation. In this case, as long as gauge-field backreaction on the inflaton field sets in after the moment of time defined through $|\varphi| \sim \varphi_{\text{thr}}$, our GW spectra are then insensitive to the choice of $V_{\text{CMB}}(\varphi)$. Let us explain this point in more detail: Around the transition in the shape of

the potential, when backreaction effects from gauge-field production are still negligible, the slow-roll evolution of the inflation field will approach the attractor solution of chaotic inflation [148, 149]. Then, for an appropriate choice of the threshold value φ_{thr} , gauge-field production will only become relevant after the system has already spent a few e -folds in the attractor solution. Hence, the subsequent inflaton evolution, gauge-field production, and SGWB spectra will have no memory of the model-dependent shape of the potential at large field values. In other words, our results derived at $|\varphi| \lesssim \varphi_{\text{thr}}$ will be identical to those of a pure $V(\phi) = m^2 \phi^2/2$ model. The conditions imposed on $V_{\text{CMB}}(\varphi)$ by our construction are minimal: continuity demands $V_{\text{CMB}}(\varphi_{\text{thr}}) = m^2 \varphi_{\text{thr}}^2/2$, and monotonicity imposes $V_{\text{CMB}}(\varphi_{\text{CMB}}) \gtrsim V_{\text{CMB}}(\varphi_{\text{thr}})$.

Following this logic, one may then constrain the mass of the inflaton, m , by using standard slow-roll results for the amplitude of scalar perturbations, A_S , and the tensor-to-scalar ratio, r as given in chapter 1,

$$A_S \simeq \frac{V}{24\pi^2 M_{\text{P}}^2 \varepsilon_V} \Big|_{\varphi=\varphi_{\text{CMB}}}, \quad r = 16\varepsilon_V \Big|_{\varphi=\varphi_{\text{CMB}}}, \quad (9.37)$$

with the slow-roll parameter $\varepsilon_V = (M_{\text{P}} V_{,\varphi}/V)^2/2$. We shall use the measured value $A_S = \exp(3.044)10^{-10} = 2.1 \times 10^{-9}$ [16, 203], and the upper bound, $r \lesssim 0.03$ [202, 203], which together imply (see also Sec. 1.5)

$$V(\varphi_{\text{thr}}) \lesssim V_{\text{CMB}}(\varphi_{\text{CMB}}) \lesssim 9.3 \times 10^{-10} M_{\text{P}}^4. \quad (9.38)$$

A second, independent constraint can be derived by considering the bound on scalar non-Gaussianity at CMB scales. This upper bound may be written as $|\xi_{\text{CMB}}| \lesssim 2.5$ [15, 49, 56, 57], which, assuming that the inflaton evolves on the slow-roll attractor at CMB scales, translates into the condition $\varepsilon_V \lesssim 12.5/\beta^2$, where ε_V is the first slow-roll parameter (cf. Eqs. (1.29), (1.33) and (3.20)). When combined with the measured value of A_S , this can be further expressed as a constraint on the inflaton potential at φ_{thr} ,

$$V(\varphi_{\text{thr}}) \lesssim V_{\text{CMB}}(\varphi_{\text{CMB}}) \lesssim 6.2 \times 10^{-6} M_{\text{P}}^4/\beta^2. \quad (9.39)$$

Next, we seek to translate these bounds on $V(\varphi_{\text{thr}})$ into bounds on the inflaton mass m . In doing so, it is worth remembering the goal of this exercise: we are interested in determining at which point the inflaton potential will have to be quadratic to a good approximation, such that GW spectra that are computed assuming a quadratic potential are indeed insensitive to the potential shape at higher scales. To do so, one can translate φ_{thr} to a characteristic threshold frequency, f_{thr} , which corresponds to the frequency of GWs today that left the horizon when $\varphi = \varphi_{\text{thr}}$. This way, GWs with frequencies above f_{thr} are sure to have originated from the inflationary dynamics evaluated in a quadratic potential. By using standard results for the dynamics of the inflaton field during chaotic inflation (see Sec. 1.2.2), one can approximately relate this threshold frequency to the inflaton potential at the threshold scale,

$$V(\varphi_{\text{thr}}) \simeq m^2 M_{\text{P}}^2 \left[51.7 - 4.6 \log_{10} \left(\frac{f_{\text{thr}}}{\text{Hz}} \right) + 2.3 \log_{10} \left(\frac{m}{M_{\text{P}}} \right) - 2\Delta N_{\text{BR}} \right]. \quad (9.40)$$

To arrive at this equation, we first use that any GW frequency, f , can be related to its corresponding comoving momentum k as $f = k/(2\pi a_0)$; then we consider that the GWs with the largest frequency originate from the end of inflation, and are determined by the Hubble rate (see Eq. (9.35)); and finally, we use that $H_{\text{end}} \simeq m/\sqrt{6}$ for chaotic inflation. To account for the possibility of strong-backreaction effects prolonging inflation past the expected slow-roll result, we also

include an additional redshift term ΔN_{BR} . While in Eq. (9.40), we assume that reheating occurs instantaneously, the effect of reheating would also be an additional source of redshift, entering into Eq. (9.40) the same way as ΔN_{BR} (cf. Eq. (4.3) in our original publication, Ref. [5]).

To see Eq. (9.40) applied, consider two scenarios (first for $\Delta N_{\text{BR}} = 0$): In a conservative construction, we may demand that the GW spectrum starting from low frequencies upwards needs to originate from the chaotic-inflation stage of Eq. (9.36). In this case, the CMB constraints from Eq. (9.38) imposes a tighter bound on the inflaton mass m , as there is less flexibility in modifying the inflaton potential at higher scales to accommodate CMB constraints. For example, choosing the threshold frequency to correspond to energy scales relevant for BBN, $f_{\text{thr}} = f_{\text{BBN}} \simeq 10^{-12}$ Hz, we find $m \lesssim 3.1 \times 10^{-6} M_{\text{P}}$. In the opposite extreme, we may state that only GW spectra at high frequencies need to originate from chaotic inflation dynamics, picking, e.g., $f_{\text{thr}} = 10^{-6}$ Hz. Then the bound by Eq. (9.38) is slightly relaxed, and we find $m \lesssim 3.8 \times 10^{-6} M_{\text{P}}$, reflecting that we gained slight flexibility in adapting the inflaton potential at higher scales. Next, consider how additional backreaction affects these bounds. Backreaction implies that inflation is prolonged without a need for the inflaton to traverse down the potential. This entails that the inflaton amplitude corresponding to f_{thr} can be lower than in a scenario without backreaction. For example, in a scenario with extreme backreaction, $\Delta N_{\text{BR}} = 20$, the bound on the inflaton mass for $f_{\text{thr}} = 10^{-6}$ Hz is relaxed; $m \lesssim 6 \times 10^{-6} M_{\text{P}}$. Note, however, that the extended duration of inflation due to backreaction should also not be too large such that it does not affect the physics corresponding to frequencies $f < f_{\text{thr}}$. This surely occurs when the term in square brackets in Eq. (9.40) becomes negative. In this discussion, we only accounted for Eq. (9.38), as the non-Gaussianity bound, Eq. (9.39), only becomes more constraining for $\beta \gtrsim 80$, and our subsequent analysis shall not extent to such large values of the dimensionless axion–vector coupling. Ultimately, we find that the bound on the inflaton mass is not very sensitive to the threshold frequency f_{thr} , thus we opt for the conservative estimate $m \lesssim 3.1 \times 10^{-6} M_{\text{P}}$ for $f_{\text{thr}} = f_{\text{BBN}}$ and $\Delta N_{\text{BR}} = 0$.

9.2.2 Gravitational-wave detection characteristics and constraints

To estimate the observational prospects for detecting a GW signal with a given GW observatory, we shall employ the signal-to-noise ratio (SNR) defined as [455–457]

$$S/N = \left(n_{\text{det}} t_{\text{obs}} \int_{f_{\text{min}}}^{f_{\text{max}}} df \left(\frac{\Omega_{\text{signal}}(f)}{\Omega_{\text{noise}}(f)} \right)^2 \right)^{1/2}. \quad (9.41)$$

The integration range $[f_{\text{min}}, f_{\text{max}}]$ corresponds to the frequency band of the detector, $n_{\text{det}} = 1$ or 2 distinguishes between auto-correlation and cross-correlation searches for an SGWB signal, and t_{obs} is the observation time. For experiments that are currently not operational, we set $t_{\text{obs}} = 1$ yr.

In our analysis, we account for three different experiments, each representative of current or next-generation detectors in their respective frequency band. For next-generation detectors, we use ET for the frequency band 1 Hz – 10 kHz [458],⁶ and LISA for the frequency band 10 μ Hz – 1 Hz [228, 460]. For the PTA frequency band, 1 nHz – 100 nHz, we estimate the potential for a GFIGW observation by the NANOGrav collaboration in their 15-year data set [242], henceforth referred to as NG15. To compute the SNR for this dataset, we work with the total observing time span covered by it, $t_{\text{obs}} = 16.03$ yrs. Additionally, we consider any sensitivity by the LIGO–Virgo network as a constraint, given that it did not observe an SGWB in its third

⁶Alternatively, one could also consider Cosmic Explorer [459], whose design sensitivity is comparable to ET, and which covers a similar frequency range.

observing run [239].⁷ We shall refer to these constraints as HLVO3 corresponding to Hanford (H), Livingston (L), and Virgo (V). These detectors cover a frequency band 10 Hz – 50 kHz [223–225]. For ET and LISA, we use the strain noise spectra Ω_{noise} collected in Ref. [461, 462]. For the HLVO3 constraints, we work with the characteristic HLV strain noise spectra available at Ref. [463]. For NG15, we take the noise spectrum Ω_{noise} provided by the NANOGrav collaboration [464, 465]. This collection of noise spectra also allows us to compute power-law-integrated sensitivity (PLIS) curves [466] for HLVO3, ET, LISA, and NG15, which we show in Figs. 9.4 and 9.8.

The energy density of GWs also contributes to the energy budget of dark radiation, which can be quantified in terms of the effective number of relativistic degrees of freedom in addition to SM photons and neutrinos. We denote the difference between the total effective number of relativistic degrees of freedom, N_{eff} , and the SM prediction $N_{\text{eff}}^{\text{SM}}$ [467] by

$$\Delta N_{\text{eff}} = N_{\text{eff}} - N_{\text{eff}}^{\text{SM}}, \quad N_{\text{eff}}^{\text{SM}} = 3.0440 \pm 0.0002, \quad (9.42)$$

which allows us to write the upper limit on the allowed amount of dark GW radiation as [219]

$$\int_{f_{\text{BBN}}}^{f_{\text{end}}} \frac{df}{f} h^2 \Omega_{\text{GW}}(f) \lesssim 5.6 \times 10^{-6} \Delta N_{\text{eff}}. \quad (9.43)$$

While Planck data alone imposes the limit $\Delta N_{\text{eff}} \lesssim 0.33$ [16], combined BBN and CMB data suffer from larger uncertainties [468, 469], which weakens this limit. Following Ref. [242], we work with a conservative limit, $\Delta N_{\text{eff}} \lesssim 0.5$ in what follows.

9.3 Gravitational waves from pure axion inflation

In this section, we now discuss the first half of our results, focussing on GFIGW production in the model of pure axion inflation (PAI). To build towards these results, we begin by considering different features of a representative benchmark GW spectrum for PAI that we may use to build intuition on how to interpret our later results. Thereafter, we proceed with our full detectability analysis. For this, we start by using analytical arguments to identify a suitable region of the (m, β) parameter space in which we can expect to find a potentially observable GW signal. We follow up on these estimates with a comprehensive parameter scan over this region and finish with a detailed discussion of our results.

9.3.1 A benchmark spectrum for pure axion inflation

In section 9.1, we have mainly discussed the production of GWs from gauge fields abstractly, noting that the resulting GW spectrum (or equivalently the tensor power spectrum) decomposes into a vacuum and an induced contribution. We now want to consider how this decomposition manifests itself in a typical GW spectrum for axion inflation when accounting for the fact that gauge-field production is a dynamical process, which starts out weak, but grows continuously as the inflaton field is rolling down its potential hill.

To discuss this, we study a representative GW spectrum which we obtain for the standard benchmark scenario $m = 6.16 \times 10^{-6} M_{\text{P}}$ and $\beta = 25$. As a reminder, this is precisely the

⁷Note that the LIGO–Virgo–KAGRA network has recently concluded their fourth observing run, and has updated their search for an SGWB by also including data from the first part of the fourth observing run of LIGO [240]. This improves upon their existing upper limits on the amplitude of an SGWB by roughly a factor of two. In what follows, we show results that are based on our original publications [4, 5], for which this dataset had not yet been available. However, given that the improved upper limit differs by an order-one factor, we expect that our subsequent results and conclusions are not qualitatively altered by these updated constraints.

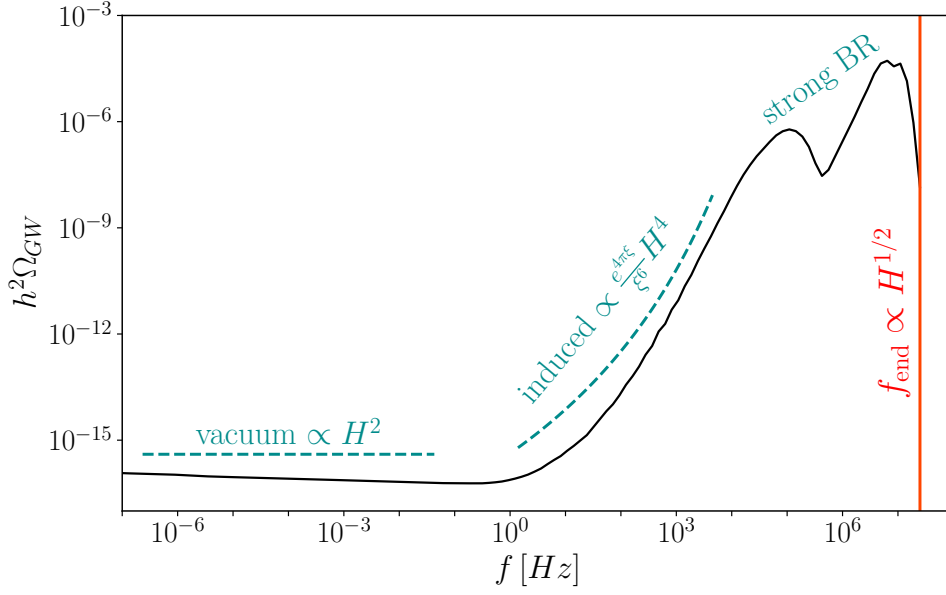


FIGURE 9.1: A representative GW spectrum for pure axion inflation which is computed for the standard benchmark $\beta = 25$ and $m = 6.16 \times 10^{-6} M_{\text{P}}$. We indicate the typical separation of the GW spectrum into three pieces. At low frequencies, vacuum GWs dominate the spectrum, then gauge-field-induced GW production takes over, finally, gauge-field production during the regime of strong backreaction can lead to additional features in Ω_{GW} .

benchmark scenario for which we previously analyzed the dynamics of axion inflation in Sec. 4.2.2 (cf. Fig. 4.1). The spectrum which we obtain for this data point is shown in Fig. 9.1. For the moment, we are only concerned with the GW spectrum itself and less so with assessing its detectability or considering CMB constraints.

To understand Fig. 9.1, we need to briefly remind ourselves of the fact that different frequencies in GWs probe different time periods during inflation. Low frequencies correspond to small comoving wavenumbers k , which have left the Hubble horizon at early times during inflation, while higher frequencies correspond to larger comoving wavenumbers, which have thus left the horizon later during inflation. Correspondingly, low frequencies probe the physics of inflation at early times, high frequencies at later times. Turning then towards the GW spectrum in Fig. 9.1, we see that it decomposes into three distinct pieces. At low frequencies, the spectrum is nearly flat, it then rises rapidly at around a frequency $f \sim 1$ Hz, and finally, it flattens out again and exhibits two local maxima. These three phases can be then mapped to the dynamical growth of gauge fields during axion inflation; at early times, gauge-field production has not yet led to a significant production of electromagnetic radiation, and thus there is no active source of GWs. The low-frequency part of the GW spectrum therefore is dominated by vacuum GWs, i.e., Ω_{GW} is primarily determined from the nearly scale-invariant $\mathcal{P}_{h,\lambda}^{\text{vac}}$ in Eq. (9.24). At some point, GFIGW production becomes efficient, and the induced contribution of the GW spectrum begins to dominate, $\mathcal{P}_{h,\lambda}^{\text{ind}} > \mathcal{P}_{h,\lambda}^{\text{vac}}$. This explains the sudden growth of the spectral amplitude in Fig. 9.1. The final stage of the spectrum then corresponds to GWs which are sourced in the strong-backreaction regime of axion inflation. The typical oscillations in the gauge-field energy densities that emerge during the strong-backreaction regime imprint themselves in the GW spectrum in the form of two peaks. The maximal frequency of the spectrum is then set by the largest wavenumber that exits the horizon before the end of inflation (cf. Eq. (9.35)).

As indicated in Fig. 9.1, we can also use simple analytical arguments to extrapolate how this

behavior changes for variations of the parameters β and m of our model. As we initialize our system in the weak-backreaction regime, the vacuum contribution to Ω_{GW} will initially always dominate. Its amplitude in the weak-backreaction regime can be estimated based on slow-roll dynamics, and scales with $H^2 \simeq m^2$ (cf. Eq. (9.26)). The induced contribution then takes over, initially for dynamics that are governed by weak backreaction, where the analytical formula for $\mathcal{P}_{h,\lambda}^{\text{ind}}$ in Eq. (9.27) approximately applies. From this, one can deduce that the induced contribution scales as $\mathcal{P}_{h,\lambda}^{\text{ind}} \propto H^4 \simeq m^4$. The coupling strength β on the other hand regulates the growth in $\mathcal{P}_{h,\lambda}^{\text{ind}}$, which scales exponentially with ξ and thus with β . For larger β , gauge-field production becomes efficient at an earlier time, which shifts the rise in the GW spectrum to lower frequencies. Furthermore, following Eq. (9.35), the overall position of the spectrum on the frequency axis is also regulated by the inflaton mass as $f_{\text{end}} \propto H^{1/2} \simeq m^{1/2}$. Finally, the contribution of Ω_{GW} from the strong-backreaction regime plays a special role. A phase of strong backreaction prolongs the duration of inflation past what is expected from slow-roll dynamics. This implies that physical frequencies are further redshifted than what would have been expected if strong-backreaction effects were absent (there is an unexpected phase of quasi-exponential growth in the scale factor). This redshift moves all features in Ω_{GW} which are not due to strong backreaction to the left. As a result, any growth in Ω_{GW} over the vacuum contribution happens already at lower frequencies than what would be expected based on purely analytical estimates assuming weak backreaction. In addition, the prolonged phase of strong backreaction allows for more gauge-field production, which further enhances the induced GW spectrum.

At this point, it is important to stress the regime of validity of our present analysis, which relies on the fact that the GEF adequately captures the inflationary dynamics of pure axion inflation. As discussed in chapter 4, this is true for the phase of weak backreaction, and the onset of the strong-backreaction regime. However, the GEF can only qualitatively resolve the nonlinear evolution of the system for strong gauge-field backreaction, because it cannot capture the dynamical growth and backreaction of inflaton inhomogeneities in this regime. Accordingly, the GW spectra we compute here are reliable up until the onset of strong backreaction, but once strong backreaction begins, we can only qualitatively capture the further sourcing of the GW spectrum, and the additional redshift in frequency. This is in contrast to lattice techniques, which can fully account for inhomogeneous backreaction.⁸

Despite this limitation, our GEF-based approach however comes with a great advantage; its cheap numerical cost makes a comprehensive analysis of GW production over a large portion of the (m, β) parameter space feasible. Lattice techniques, which can capture the full phase of inhomogeneous backreaction, are limited to the study of a few benchmark points. To our knowledge, there is only a single lattice study which considers GFIGW production during inflation, namely Ref. [380], and their analysis is limited to four benchmark spectra for different values of m and β . A second advantage of our GEF-based approach is that we can also capture the dynamical evolution of the inflationary system over more than 60 e -foldings, thus many orders of magnitude in GW frequencies. This is in contrast to lattice results, whose dynamical range is more limited. Considering that one needs to capture both the growth of the induced contribution and the subsequent evolution of the spectrum over multiple orders of magnitude in frequency, the GEF approach is well suited for the type of analysis we perform here. The way to therefore think of our GEF results is that they establish a *GEF benchmark*, a type of precursor analysis to identify regions of parameter space which warrant further exploration using lattice techniques.

We should also mention that our results rely on the assumption that GFIGW production ends once inflation has ended. This is a simplification, as gauge-field production, and thus the sourcing

⁸One can expect that in this regime, also inflaton inhomogeneities may become important as a source of GWs, as the source term Σ_ϕ in Eq. (9.5) can become relevant in this regime.

of GWs, can continue into a stage of preheating. To reliably compute this effect, however, also necessitates lattice simulations. Such an analysis has been performed in a set of two papers, Refs. [385, 386] for two fixed values of the inflaton mass m .⁹ In what follows, we shall use the results in Refs. [385, 386] as an indication for the importance of preheating effects, and to guide our conclusions.¹⁰

With this first intuitive understanding of GFIGW production in place, and a keen awareness of the limitations of our approach, we are then fully prepared to discuss our results of our extensive parameter-space analysis.

9.3.2 Identifying relevant parameter-space regions

To begin, we first aim to identify a relevant region of parameter space where GW production is not so weak as to not be observable, while not being so strong that it violates the bound $\Delta N_{\text{eff}} < 0.5$ in Eq. (9.43). To identify the conditions that are needed to achieve this, we estimate the expected GW amplitude by relying on the approximate analytical formula for the induced tensor power spectrum in Eq. (9.27). This strategy directly follows from our qualitative analysis of the benchmark spectrum in Fig. 9.1.

Considering that the induced tensor power spectrum in the weak-backreaction regime scales with H and ξ as in Eq. (9.27), we can determine how a decrease in the inflaton mass $m \propto H$ needs to be compensated by an increase in the axion–vector coupling β to maintain an equal amplitude in the tensor power spectrum. Using Eq. (9.27), and taking into account that $|\xi| = \beta|\partial\varphi/\partial N|/(2M_{\text{P}})$, one arrives at

$$\log_{10} \left(\frac{m}{M_{\text{P}}} \right) \simeq -\frac{\pi}{2\log(10)} \left| \frac{1}{M_{\text{P}}} \frac{\partial\varphi}{\partial N} \right| \beta + C, \quad (9.44)$$

with C an arbitrary constant. For chaotic inflation, we have that $|\partial\varphi/\partial N| \lesssim 1.4M_{\text{P}}$ (cf. Eq. (1.34)), which we can use with the analytical expression for $\mathcal{P}_{h,\lambda}$ in Eq. (9.27) to estimate that

$$\log_{10}(h^2\Omega_{\text{GW}}^{\text{min}})^{1/4} + 3.3 \lesssim C \lesssim \log_{10}(h^2\Omega_{\text{GW}}^{\text{max}})^{1/4} + 3.3, \quad (9.45)$$

where $h^2\Omega_{\text{GW}}^{\text{min}}$ and $h^2\Omega_{\text{GW}}^{\text{max}}$ are the minimal and maximal amplitudes in between which the SGWB signal is supposed to fall. To expect any observable signal for LISA or ET (whose expected sensitivity reach extends to the lowest values of $h^2\Omega_{\text{GW}}$), we conservatively estimate $h^2\Omega_{\text{GW}}^{\text{min}} \approx 10^{-15}$, allowing for strong-backreaction effects to enhance the signal after the slow-roll phase. At the same time, we estimate $h^2\Omega_{\text{GW}}^{\text{max}} \approx 10^{-5}$ to ensure $\Delta N_{\text{eff}} < 0.5$. Collectively, this implies that observable GW signals should approximately be attainable for m and β following a log-linear relationship as in Eq. (9.44) with the constant offset ranging from $-0.45 \lesssim C \lesssim 2.05$.

We then perform a first preliminary scan of the (m, β) parameter space to further refine these expectations. In this first parameter scan, we identify that parameter points which are excluded by $\Delta N_{\text{eff}} > 0.5$ all lie above a line following the log-linear relationship $\log_{10}(m/M_{\text{P}}) \gtrsim \tan\theta\beta + 0.3$ with $\tan\theta \approx -0.4$. This motivates us to work in a suitably rotated parameter space described by the two parameters

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \beta \\ \log_{10}(m/M_{\text{P}}) \end{pmatrix} \quad (9.46)$$

⁹See also Ref. [85], the authors of which also comment on the necessity for capturing GFIGW production into the reheating phase.

¹⁰Note that these authors used $\Delta N_{\text{eff}} \lesssim 0.33$ when assessing the possibility for ΔN_{eff} violation for GWs from preheating, which is directly inferred from the Planck result for N_{eff} , Ref. [16]. This is in contrast to our work, which uses $\Delta N_{\text{eff}} \lesssim 0.5$, to better reflect the uncertainties regarding BBN determinations of N_{eff} , as explained previously. This small difference in the upper limit on ΔN_{eff} however proves insignificant for our later discussion.

with $\cos \theta \approx 0.93$, $\sin \theta \approx -0.37$. Variations of y thus correspond to varying the constant offset of the line $\log_{10}(m/M_P) = \tan \theta \beta + C$, with $y = \cos \theta C \gtrsim -0.44$. We then perform a refined parameter space scan over 51×21 evenly spaced points $y \in [0.15, 0.55]$ for $x \in [10, 60]$. This narrow range in y coordinates is chosen such that we can accurately determine the line of $\Delta N_{\text{eff}} = 0.5$, while also allowing us to identify the regions where GW signals are potentially observable with a signal-to-noise ratio larger than one.

9.3.3 Results for pure axion inflation

The results of our parameter scan can be summarized as follows: we find that ΔN_{eff} excludes a large region of the (m, β) parameter space, constraining regions which otherwise would be allowed by CMB bounds on the inflaton mass m based on the tensor-to-scalar ratio in Eq. (9.38) or the non-Gaussianity bound in Eq. (9.39). Moreover, for ET, LISA, and HLVO3, we do *not* find a single parameter space point with a signal-to-noise ratio exceeding unity which is not simultaneously in conflict with ΔN_{eff} .¹¹ To elaborate upon these findings, we show the results of our analysis in Figs. 9.2 and 9.3, once in terms of the model parameters β and m (Fig. 9.2), and once in terms of x and y (Fig. 9.3).

Let us begin with Fig. 9.2. Here, we show the region which is excluded by $\Delta N_{\text{eff}} > 0.5$ as a gray shaded area, and CMB constraints on the inflaton mass following from Eqs. (9.38) and (9.39) in purple. The CMB constraints are placed at $m \lesssim 3.1 \times 10^{-6} M_P$ which follows from Eq. (9.40) and Eq. (9.38) for $f_{\text{thr}} = f_{\text{BBN}}$ and $\Delta N_{\text{BR}} = 0$. We may immediately note that the bound coming from ΔN_{eff} is evidently more stringent than CMB bounds on m , constraining far lower inflaton masses for larger axion–vector couplings β . In addition to these bounds, we also highlight three regions in which we find a signal-to-noise ratio above one for the three experiments ET, LISA and HLVO3 in red, green and blue, respectively. All regions of parameter space wherein a GW signal would be potentially observable lie above the line excluded by ΔN_{eff} , putting into question their phenomenological viability.

As a point of comparison, we also indicate the results of Refs. [380, 386] in Fig. 9.2. The two points marked by stars give the results of Ref. [386] with $\beta \simeq 14$, $m \simeq 6.16 \times 10^{-6} M_P$ (red) and $\beta \simeq 16$, $m \simeq 6.16 \times 10^{-7} M_P$ (yellow). For these points, Ref. [386] found violations of $\Delta N_{\text{eff}} < 0.33$ by accounting for gauge-field production during preheating. The β values of these points are estimated based on Fig. 1 in Ref. [386] and may not be exact. As these points lie *below* our excluded region, they emphasize that post-inflationary GFIGW production may be more important than what our methods can account for. The results marked by dots indicate the benchmark points studied in Ref. [380]. These are $\beta \simeq 10, 12, 15$, and 18 for $m \simeq 5.31 \times 10^{-6} M_P$, which we respectively mark in light-blue, dark-blue, orange, and red. For the first two points, Ref. [380] found ΔN_{eff} -violation from gauge-field production after inflation (in accordance with Ref. [386]). For the latter two, they instead obtained no ΔN_{eff} -violation during inflation, which already accounts for strong backreaction including inflaton inhomogeneities. These lattice data points clearly underpin that the excluded region, $\Delta N_{\text{eff}} > 0.5$, following from our results is not a hard bound on PAI. Rather, as previously explained, our results are a GEF benchmark meant to guide future lattice studies that can go beyond the limitations of our analysis. Nevertheless, as we shall discuss in the subsequent paragraphs, our results still allow us to draw important conclusions about the

¹¹We also do not find a region with $S/N > 1$ for NG15, but this can be put down to the region of parameter space we have explored. An observable signal at PTA frequencies requires a GW spectrum which rises already at $f \sim 10^{-9}$ Hz over the vacuum contribution. Such a swift rise in the GW spectrum requires larger values of β than what we captured in our parameter-space scan. In particular, extrapolation suggests that $\beta \sim 100$ is necessary (see also table 9.1), but technical limitations prohibit us from scanning above $\beta \gtrsim 60$. For similar reasons, we also do not capture the full LISA frequency band in our analysis, which also would require β values as large as 80.

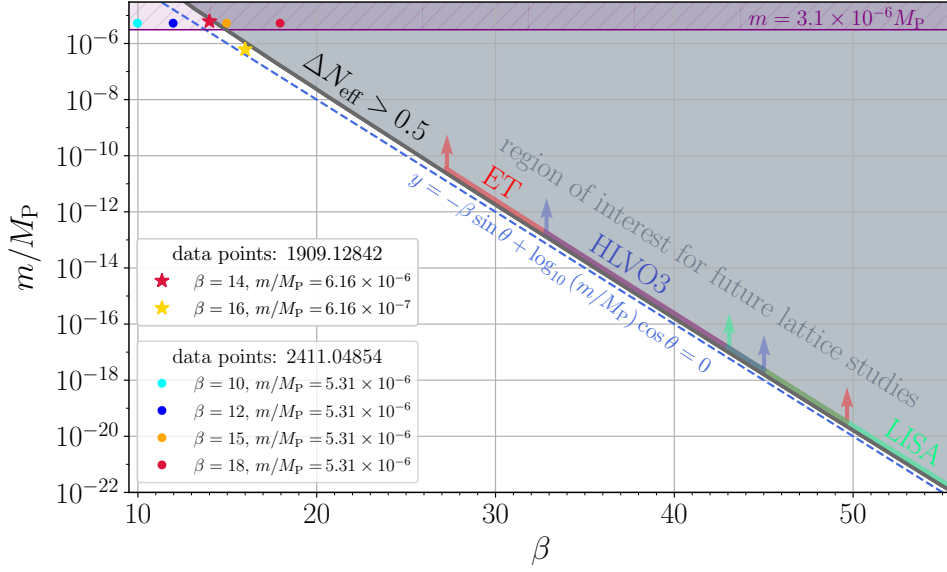


FIGURE 9.2: Results of our final parameter scan for PAI shown in terms of β and m/M_P . The gray-shaded area shows the region where $\Delta N_{\text{eff}} > 0.5$, implying an overproduction of GFIGWs in conflict with CMB and BBN constraints. The purple-shaded area shows $m > 3.1 \times 10^{-6} M_P$, which is excluded due to the bound on the tensor-to-scalar ratio as given by Eq. (9.38) for a threshold frequency $f_{\text{thr}} = f_{\text{BBN}}$ and $\Delta N_{\text{BR}} = 0$. We also depict regions of $S/N > 1$ for ET (red), LISA (green), and HLVO3 (red). The red and yellow stars, respectively, indicate $\beta = 14$, $m = 6.16 \times 10^{-6} M_P$ and $\beta = 16$, $m = 6.16 \times 10^{-7} M_P$, the approximate parameter points for which Ref. [386] found $\Delta N_{\text{eff}} > 0.33$ for gauge-field production during preheating. The light-blue, blue, orange, and red points indicate $\beta \simeq 10, 12, 15$, and 18 for $m \simeq 5.31 \times 10^{-6}$ corresponding to results obtained by Ref. [380]. The first two points violate ΔN_{eff} bounds only during preheating, the latter two avoid ΔN_{eff} bounds due to inhomogeneous-backreaction effects (cf. Sec. 4.3).

onset of strong backreaction and the detection prospects for GFIGWs from PAI in next-generation interferometer experiments.

In Fig. 9.3, we show our results in the rotated (x, y) coordinate plane, which allows us to better resolve several interesting features of our results. In dashed black lines, we indicate reference lines of constant β and m/M_P . In this very narrow diagonal slice of the β - $\log_{10}(m/M_P)$ plane, variations in y along constant x keep β roughly constant while only very mildly varying m/M_P .

As in Fig. 9.2, the area with $\Delta N_{\text{eff}} > 0.5$ is shown as shaded in gray and the one with $m > 3.1 \times 10^{-6} M_P$ is shaded in purple. We also indicate regions of $S/N > 1$ shaded in red, blue, and green for ET, HLVO3, and LISA, respectively. By this, we can see more clearly that there are no regions of $S/N > 1$ for these experiments which also lie outside the region constrained by ΔN_{eff} . The green- and red-shaded regions at $x \lesssim 12$ correspond to points for which $\mathcal{P}_h^{\text{vac}}$ would be observable with LISA and ET. These regions are, however, excluded by our CMB bounds. Furthermore, in dashed white lines, we indicate regions where the duration of inflation is extended by more than ΔN_{BR} e -folds — a smoking-gun signature of strong backreaction. Notably, $\Delta N_{\text{BR}} > 3, 4, 5$ always occurs simultaneously to $\Delta N_{\text{eff}} > 0.5$. This clearly suggests a direct correspondence between ΔN_{eff} -violation and strong backreaction. A clear interpretation is that the extended period of gauge-field production during the phase of strong backreaction leads to the observed overproduction of gravitational radiation. This interpretation is strengthened by the fact that regions where backreaction effects are absent (indicated by the contour line for $\Delta N_{\text{BR}} < 1$ in dashed brown), lie clearly outside the boundary describing $\Delta N_{\text{eff}} > 0.5$. Apparently, when strong

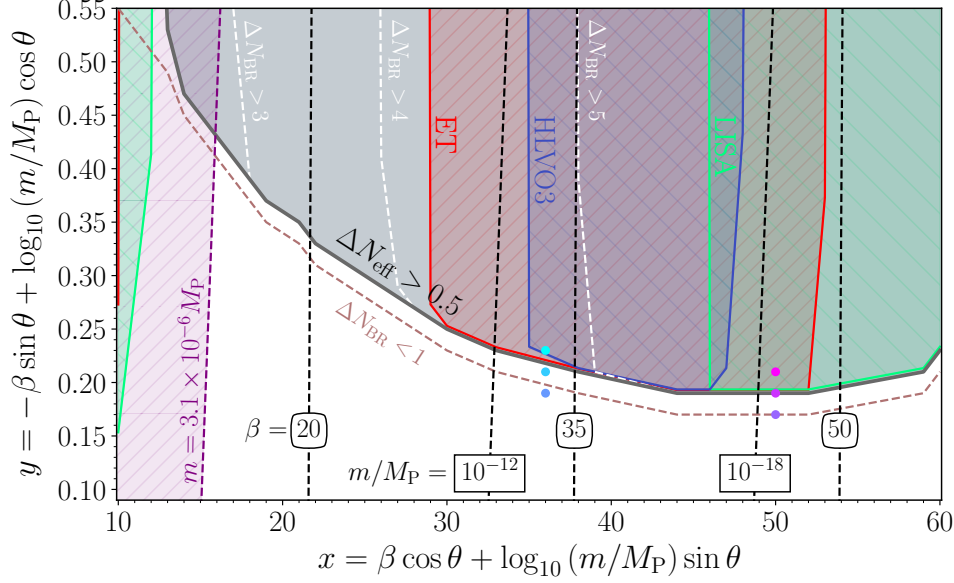


FIGURE 9.3: Same results as in Fig. 9.2 shown in terms of x and y . As before, the gray-shaded area indicates $\Delta N_{\text{eff}} > 0.5$ and the purple-shaded area indicates $m > 3.1 \times 10^{-6}$. The red, blue, and green-shaded areas indicate $S/N > 1$ for ET, HLVO3, and LISA, respectively. Clearly, $S/N > 1$ is only possible for parameter-space points in conflict with $\Delta N_{\text{eff}} > 0.5$. In white dashed lines, we indicate where the duration of inflation is extended by more than $\Delta N_{\text{BR}} \simeq 3, 4, 5$ due to strong-backreaction effects. In dashed brown, we show where the extended duration of inflation is less than one e -fold, $\Delta N_{\text{BR}} < 1$. Evidently, strong backreaction is required for an observable signal, but simultaneously leads to an overproduction of gravitational radiation. The points indicated in shades of light-blue are $x = 36$ and $y = 0.19, 0.21, 0.23$. The magenta points are $x = 50$ and $y = 0.17, 0.19, 0.21$. SGWB spectra for these six benchmark points are shown in Fig 9.4.

backreaction does not take place, ΔN_{eff} -violation can be avoided, but the resulting GW spectra also fall short of the sensitivity region for ET, LISA, or HLVO3.

To expand upon this observation, we select six benchmark points for which we show the full SGWB spectra in Fig. 9.4. These points are indicated in Fig. 9.3 by three light-blue or magenta points that, respectively, refer to $x = 36, y = 0.19, 0.21, 0.23$ or $x = 50, y = 0.17, 0.19, 0.21$. The same color scheme is also used for the corresponding GW spectra in Fig. 9.4. What can be seen by these spectra is that a marginal increase in the inflaton mass, m , ($\simeq y$) for a fixed value of β can lead to a sudden and drastic increase in the amplitude of Ω_{GW} . Although this amplification is needed to reach the sensitivity range of LISA, ET or HLVO3, it is too violent and leads to a conflict with ΔN_{eff} . Comparing to Fig. 9.3, it is notable that this sudden increase occurs only for benchmark points which undergo a prolonged phase of expansion, which therefore indicates that the observed amplification of the GW spectra can be attributed to strong-backreaction effects.¹²

The physical process behind this can be explained as follows. Below a certain threshold mass, (e.g., $m \simeq 6.7 \times 10^{-14} M_{\text{P}}$ for $\beta = 33.5$ or $m \simeq 3.9 \times 10^{-19} M_{\text{P}}$ for $\beta = 46.5$), gauge-field-induced friction on the inflaton is just small enough such that inflation can still end on the usual slow-roll trajectory, i.e., reaching $\dot{\varphi}^2 \simeq V(\varphi)$. However, only a very mild increase past this threshold mass allows a marginally larger amount of gauge-field production, resulting in a slight increase of gauge-field friction. This prevents the inflaton field from accelerating enough to end inflation, and the system enters the strong-backreaction regime. Once this process is triggered,

¹²Comparing these spectra to Fig. 9.1, the reader may wonder at the absence of the flat vacuum part of the GW spectrum. The vacuum contribution for these spectra is strongly suppressed given the low inflaton mass m , see Eq. (9.24).

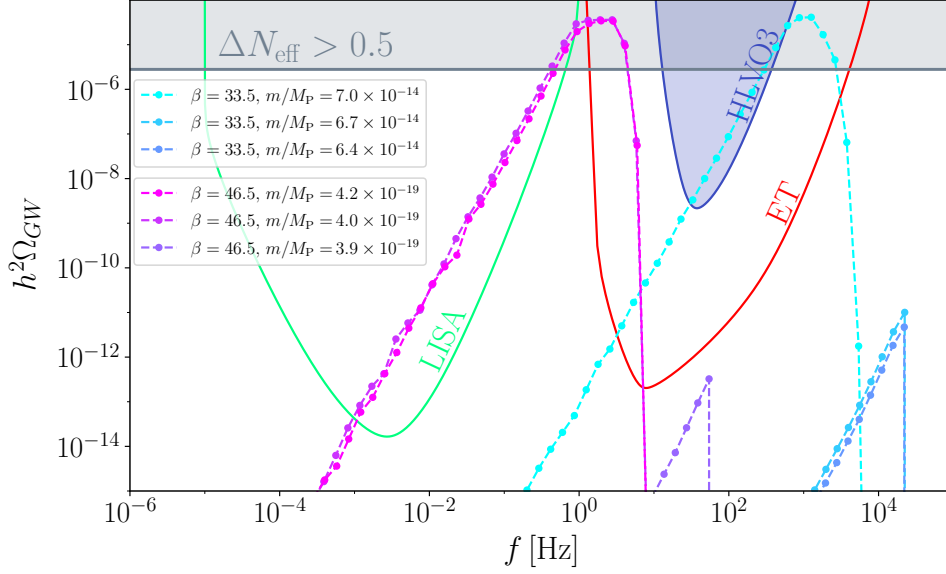


FIGURE 9.4: SGWB spectra for the six benchmark points marked in Fig. 9.3. We explicitly plot our numerical data points and connect them by piece-wise linear interpolation. Spectra in shades of light-blue correspond to $\beta = 33.5$ and $m/M_P = 6.4, 6.7, 7.0 \times 10^{-14} M_P$ or $x = 36$ and $y = 0.19, 0.21, 0.23$, those in shades of magenta correspond to $\beta = 46.5$ and $m = 3.9, 4.0, 4.2 \times 10^{-19} M_P$ or $x = 50$ and $y = 0.17, 0.19, 0.21$. In red, blue, and green, we show PLIS curves for ET, LISA and HLV03. In gray, we suggest $\Delta N_{\text{eff}} > 0.5$ (although Eq. (9.43) is an integrated bound). For each set of three benchmark points, one can notice a qualitative threshold mass, $m \simeq 6.7 \times 10^{-14} M_P$ for $\beta = 33.5$ or $m \simeq 3.9 \times 10^{-19} M_P$ for $\beta = 46.5$. A small increase past this threshold mass triggers strong backreaction, leading to a drastically increased SGWB amplitude, while also redshifting the signal to lower frequencies. Without strong backreaction, signals are too weak to be observed, with strong backreaction, the signal conflicts with $\Delta N_{\text{eff}} < 0.5$.

the end of inflation is delayed by multiple e -folds as the inflaton's kinetic energy oscillates and the energy density in the electromagnetic field further increases. During this stage, the additional growth in electromagnetic radiation further sources GWs, a process which is only interrupted once the electromagnetic energy density becomes dominant, $\rho_{\text{EM}} \simeq V(\varphi)$. By the time inflation has ended this way, the amplitude in GWs has been increased by several orders of magnitude above the levels before strong backreaction had set in.

To cement this interpretation, it is instructive to study the behavior of the backreaction parameter δ_{KG} for these benchmark points (cf. Eq. (3.27)). As a reminder, δ_{KG} quantifies the importance of gauge-field friction, with $\delta_{\text{KG}} \sim 1$ signaling strong backreaction. In Fig. 9.5, we show the evolution of δ_{KG} as a function of e -folds just before the expected end of slow-roll inflation. The parameter space points are the same as in Fig. 9.4, and we use a consistent color scheme. For the three parameter points that manage to avoid $\Delta N_{\text{eff}} > 0.5$, inflation ends on the slow-roll attractor ($\Delta N_{\text{BR}} \simeq 0$) just before $\delta_{\text{KG}} = 1$, thus never triggering strong backreaction. The opposite is true for the points where the ΔN_{eff} -bound is violated. To explain this observation, we can note the approximate scaling $\delta_{\text{KG}} \propto H^2 \propto m^2$ that is suggested by Eq. (3.32) for as long as weak backreaction persists. This entails that a marginally larger inflaton mass implies a marginally increased relevance of gauge-field friction effects. As a consequence, for slightly larger m , the inflaton field is slightly more sensitive to gauge-field friction, keeping the inflaton from reaching the end of slow-roll inflation by a fraction of an e -fold longer. This slight increase in the duration of inflation also leaves the potential for gauge-field production to continue marginally longer, buying

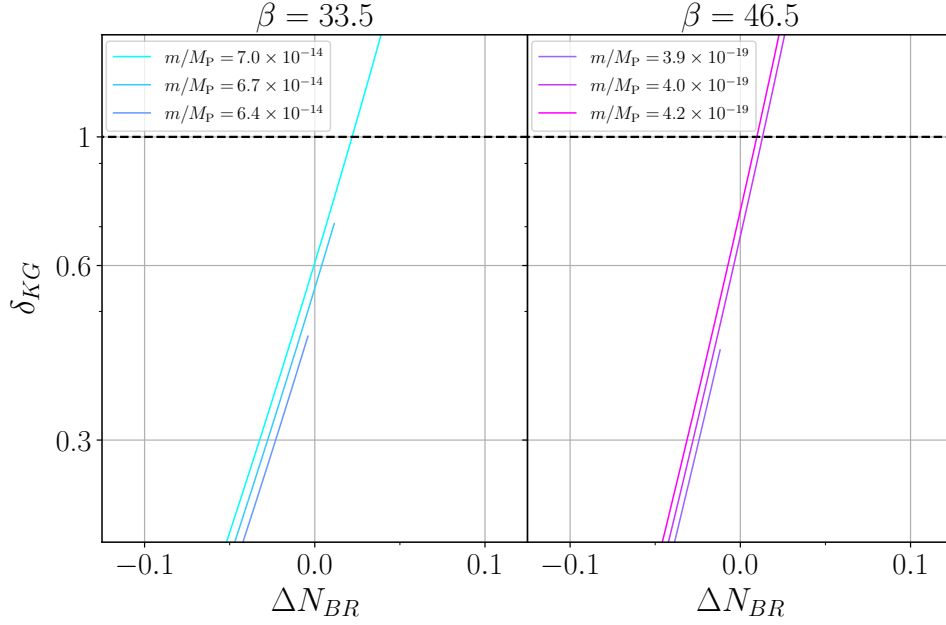


FIGURE 9.5: Evolution of the backreaction parameter δ_{KG} for the six benchmark points shown in Fig. 9.4 with the corresponding color scheme. On the horizontal axis, we show the number of e -folds ΔN_{BR} after the naive end of slow-roll inflation, approximately 61 e -folds after initializing our system. Notably, for the benchmark points with $\Delta N_{eff} < 0.5$, inflation ends just before $\delta_{KG} = 1$.

just enough time to tip the system over the edge of strong backreaction.

9.3.4 Summary on constraints and detectability regions

Following this discussion, we can state that our results entail two important observations: first, we find that strong backreaction appears to be a prerequisite for an observable GW signal from PAI for both LISA and ET. Moreover, we observe that strong-backreaction effects inevitably lead to a conflict with ΔN_{eff} bounds. Together, the two results imply that there is a limited potential for observing GWs in our model of pure axion inflation. Both of these observations, however, warrant further discussion.

Let us first focus on the bounds implied by ΔN_{eff} . We can summarize this bound by performing an empirical fit to the boundary of the region shown in Fig. 9.3. Motivated by the approximate scaling relation between β and $\log_{10}(m/M_P)$ in Eq. (9.44), we describe the viable parameter region by

$$\log_{10} \left(\frac{m}{M_P} \right) \lesssim t_1 \beta + t_2 \log_{10} \left(\frac{\beta}{10} \right) + C, \quad (9.47)$$

allowing for a mild deviation from a linear relation to capture logarithmic terms neglected in the derivation of Eq. (9.44). Our fit yields the coefficients $t_1 = -0.387$, $t_2 = -1.34$, and $C = 0.51$. Note that t_1 is consistent with the initial parameter scan, $\tan \theta \simeq -0.4 \simeq t_1$. Within the regime of validity of our approach, points which lie above this empirical contour region are ruled out by an overproduction of GWs.

However, it should be clearly stated that this line does not represent a hard exclusion limit, as is made clear by the results of Refs. [380, 385, 386], which we have indicated in Fig. 9.2. Our results

Detector	Parameter scan		Extrapolation	
	β	m/M_{P}	β	m/M_{P}
HLVO3	32.5 – 45.0	$3 \times 10^{-18} - 2 \times 10^{-13}$	—	
ET	27.0 – 49.5	$5 \times 10^{-20} - 5 \times 10^{-11}$	—	
LISA	42.5 – 56.0	$8 \times 10^{-23} - 2 \times 10^{-17}$	$\lesssim 82.0$	$\gtrsim 5 \times 10^{-33}$
NANOGrav	—		$\lesssim 107.0$	$\gtrsim 5 \times 10^{-43}$

TABLE 9.1: Regions of interest for future studies including inhomogeneous backreaction.

are limited by our computational method, as we are not able to account for inhomogeneous-backreaction dynamics nor can we consistently follow gauge-field production into the post-inflationary epoch to study GW production during preheating. References [385, 386] have shown that, for $\beta = 14$, $m = 6.16 \times 10^{-6} M_{\text{P}}$ and $\beta = 16$, $m = 6.16 \times 10^{-7} M_{\text{P}}$, gauge-field production leads to a post-inflationary violation of ΔN_{eff} . The same was found by Ref. [380] for $\beta = 10, 12$ and $m = 5.31 \times 10^{-6} M_{\text{P}}$. In all these four cases, their parameter points lie below our excluded region. From this observation, we may deduce that an absence of gauge-field backreaction during inflation does not necessarily imply that gauge-field dynamics cannot be enhanced by such effects after inflation has ended.

However, our results do appear consistent within the scope of our method: For $\beta = 10, 12$ and $m = 5.31 \times 10^{-6} M_{\text{P}}$, the benchmark points studied in Ref. [380], the authors find that strong backreaction is not reached during inflation, which is in agreement with our results. This accordance is consistent with the direct comparison between lattice and GEF results in Ref. [378, 379] (see also Fig. 4.3), which demonstrates that GEF results accurately capture the weak-backreaction regime as well as the onset of strong backreaction. This consistency clearly suggests that the first major result of our analysis is consistent in the lead up to the strong-backreaction regime: observable GW signals from PAI within the reach of current or near-future interferometers can only be attained for parameter values that also result in strong backreaction.

Meanwhile, the onset of strong backreaction may not necessarily imply an overproduction of gauge-field-sourced gravitational radiation, unlike what our results may imply. Comparing again to the lattice simulation performed in Ref. [380], we find that their data points $\beta = 15, 18$ and $m = 5.31 \times 10^{-6} M_{\text{P}}$ are also consistent with an extended period of inflation due to backreaction effects, while avoiding $\Delta N_{\text{eff}} > 0.5$, as the coupling of gauge fields to inflaton gradients dampens gauge-field production, and thus the creation of GFIGWs. In this way, their results manage to penetrate our excluded region indicated by Eq. (9.47). However, given the dynamical range covered in their simulations, it is not evident if accounting for the dynamics after inflation may lead to an overproduction of GFIGWs, after all.

These caveats notwithstanding, the relevance of the contour line in Eq. (9.47) should not be understated. Unlike the lattice results discussed above, the GEF allows us to perform an extensive parameter scan of the PAI parameter space, which allows us to draw conclusions beyond the study of specific benchmark points. The exclusion region in Eq. (9.47) represents a GEF benchmark that can be used to guide follow-up lattice studies, indicating regions of interest wherein potentially observable signals for current or future GW observatories may lie. We summarize these regions in Table 9.1. The ranges indicated in this table should be understood as lying above the contour given by Eq. (9.47) (see also Fig. 9.2). For homogeneous backreaction, we find that all regions of interest within the range of our parameter scan are in conflict with $\Delta N_{\text{eff}} > 0.5$. Solving the full dynamics of PAI including inhomogeneous backreaction, parts of these regions may yield

observable (excludable in the case of HLVO3) signals. For LISA and NANOGrav, we extrapolate the regions of interest based on Eq. (9.47) and the frequency corresponding to the end of inflation estimated in Eq. (9.35).

9.4 Gravitational waves from fermionic axion inflation

We now follow up on our results for the model of PAI by studying GW production in the model of FAI, particularly for the case where the Schwinger effect encompasses SM fermion production, i.e., FAI. As for PAI, we begin our discussion by considering a single benchmark spectrum for FAI to build an expectation for our analysis and to highlight the differences with respect to the model discussed in the previous section. We then discuss the results of our full GW analysis based on a comprehensive parameter-space exploration for FAI.

9.4.1 A benchmark spectrum for fermionic axion inflation

The main difference between PAI and FAI regarding the sourcing of GFIGWs is the fact that Schwinger pair creation inhibits the production of electromagnetic fields (cf. chapter 6). This reduction in the efficiency of gauge-field production directly implies that there is less electromagnetic radiation which can source GWs in the case of FAI. To illustrate how this effect manifests itself in the GW spectrum, Fig. 9.6 shows two GW spectra computed for the same parameter space point, $\beta = 30$ and $m = 6.16 \times 10^{-6} M_{\text{P}}$, once for the PAI model and once for the FAI model. For this first comparison, we are at this stage not concerned with experimental constraints, and merely consider the qualitative differences between the two spectra.

Looking at Fig 9.6, we can see that the amplitude in the FAI spectrum (magenta) is suppressed by several orders of magnitude with respect to the PAI spectrum (teal). Clearly, this effect can be attributed to Schwinger damping, as merely the amplitude of the gauge-field-induced part of the spectrum is lessened, while the flat vacuum part of the spectrum remains unaffected. Similarly, this explains why the GW spectrum for FAI rises more gradually than the PAI spectrum. The second noticeable feature is the absence of peaked features at high frequencies in the case of FAI. This also can be attributed to the reduced efficiency of gauge-field production for FAI, as this hinders the onset of the strong-backreaction dynamics that are otherwise present for PAI. The same effect also explains why the induced contribution rises at mildly higher frequencies for FAI than it does for PAI, as the prolonged duration of inflation associated with backreaction shifts features in the GW spectrum to the left, as explained in Sec. 9.3.1.

Based on these considerations, we need to revisit some of our previous estimates for the relevant parameter space region for our GW analysis of FAI. Perhaps surprisingly at first glance, our parameter scan needs to be performed over a smaller range in inflaton masses m , and not primarily larger couplings β . The reason is that the induced contribution to the tensor power spectrum, $\mathcal{P}_{h,\lambda}^{\text{ind}}$ still scales as $H^4 \simeq m^4$, even for the model of FAI. Although the analytical formula for $\mathcal{P}_{h,\lambda}^{\text{ind}}$ in Eq. (9.27) is no longer valid, the scaling with H in fact already follows from the full integral expression for this quantity, Eq. (9.25). The inhibited sourcing of GFIGWs in FAI implies that gauge-field spectra grow less rapidly, thus to achieve an appreciable amplitude in GWs, we need to consider large values of m . Lowering m can no longer be compensated as efficiently by an increase in β , which has a reduced effect regarding the amplification of the spectrum compared to the case of PAI. Consequently, the parameter-space analysis we perform in what follows is not performed in rotated coordinates as in the case of PAI (see Eq. (9.46)), but rather it proves prudent to perform a linear scan in β and $\log_{10} \frac{m}{M_{\text{P}}}$.

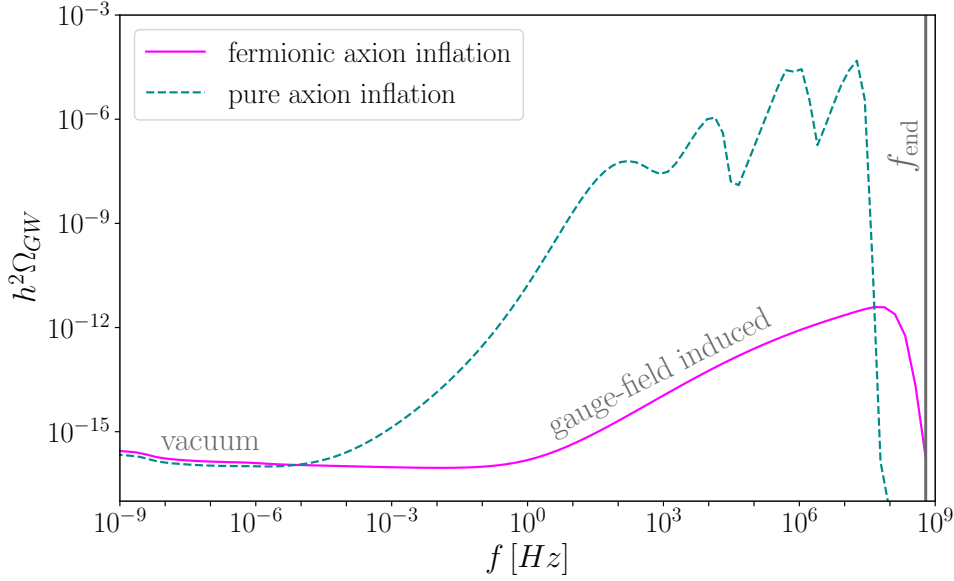


FIGURE 9.6: A representative GW spectrum for FAI (magenta) compared to one for PAI (teal). Both are computed for the benchmark $\beta = 30$ and $m = 6.16 \times 10^{-6} M_{\text{P}}$. The amplitude for the gauge-field-induced part of the FAI spectrum is reduced because of the damping effects of fermions. Similarly, there are no additional peaks at high frequencies, as gauge-field backreaction remains inefficient.

Since our numerical analysis of GFIGW production for FAI is the first of its kind, we do not have a basis for comparing our results to the literature. In particular, lacking any lattice analyses in the region of parameter space we shall focus on, we have no basis on which to assess the importance of inflaton inhomogeneities for our results. Following our discussion in chapters 4, we can however speculate that inflaton inhomogeneities are unimportant for FAI, at least as long as strong-backreaction effects are irrelevant. Similarly, there is no study of GW production during preheating for FAI.

9.4.2 Results for gauge-field-induced gravitational waves

For our study of GFIGW production in the model of FAI, we perform a parameter scan over inflaton masses ranging between $10^{-7.5} - 10^{-5.5} M_{\text{P}}$ and β between $20 - 60$. The main results of this analysis are summarized in Fig. 9.7. Here we indicate regions with $S/N > 1$ for ET (red), LISA (green), and NG15 (orange), with CMB constraints on the inflaton mass, $m \lesssim 3.1 \times 10^{-6} M_{\text{P}}$, given in purple. Evidently, for $\beta \gtrsim 30$, FAI supports sufficiently strong gauge-field production such that the resulting GWs would be detectable by both LISA and ET. In fact, we find a large overlap between regions of $S/N > 1$ for ET and LISA, indicating that the GW amplitude spans a large frequency range. For $\beta \gtrsim 54$ and $m \gtrsim 2 \times 10^{-5} M_{\text{P}}$, the signal can even extend down to the nanohertz regime where the GW signal could potentially explain the evidence for an SGWB in the NG15 dataset. However, to support large enough amplitudes in Ω_{GW} , the inflaton mass needs to be sufficiently high seeing as gauge-field production is less efficient in the presence of fermions. The need for large inflaton masses entails that a large part of the LISA and ET region of interest is in tension with the CMB bound on m , but a small viable region remains for $\beta \gtrsim 45$. The region of interest for NG15 is clearly excluded by the CMB bound. In contrast, we find that neither ΔN_{eff} nor HLVO3 impose any constraint in the region of parameter space that we analyze.

Alongside the SNR regions in Fig 9.7, we also indicate the relevance of gauge-field backreaction by showing contour lines for the extended duration of inflation due to backreaction, ΔN_{BR} . These

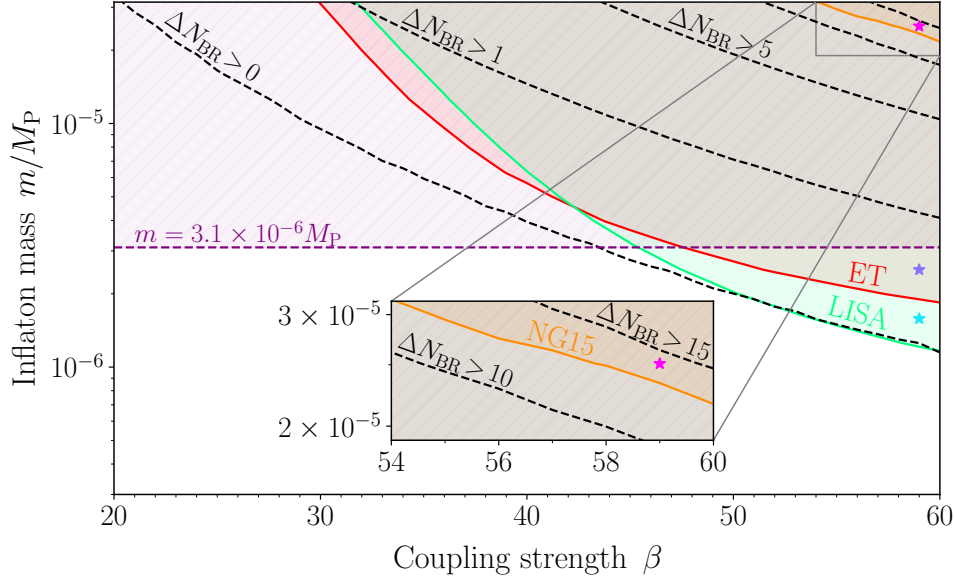


FIGURE 9.7: Regions with $S/N > 1$ for ET (red), LISA (green), and NG15 (orange) for FAI. Planck constraints on the inflaton mass are shown as a purple-shaded region. In dashed black, we show contour lines for ΔN_{BR} , the prolonged duration of inflation due to gauge-field backreaction. The benchmark points marked in cyan, violet, and magenta correspond to the GW spectra shown in Fig 9.8. For $\beta \gtrsim 45$, we find a viable region for a detection by ET or LISA, while a FAI interpretation of the NG15 signal appears to be excluded by Planck data.

contours match with those shown in Fig. 6.7, where we first discussed the regime of fermion-tempered backreaction (cf. Sec. 6.4). As we have noted as part of this discussion, larger masses and couplings increase the relevance of backreaction. However, the implications of strong backreaction for our present GW analysis are in stark contrast to our findings for PAI in the previous section. Firstly, we note that, for FAI, strong backreaction is not required to obtain an observable GW signal. Moreover, strong backreaction is not inevitably tied to a violation of $\Delta N_{\text{eff}} \lesssim 0.5$ in this model. In particular, ΔN_{eff} appears to be largely irrelevant for FAI, even for an extremely increased duration of inflation, $\Delta N_{\text{BR}} \gtrsim 15$. This can be clearly attributed to the distinct dynamical regime of fermion-tempered backreaction that we discussed in Sec. 6.4, as it allows for a delayed end of inflation without being accompanied by an overproduction of gauge bosons.

Also shown in Fig 9.7 are three benchmark points for $\beta = 59$ with $m = 1.5 \times 10^{-6} M_{\text{P}}$, $2.5 \times 10^{-6} M_{\text{P}}$ and $2.5 \times 10^{-5} M_{\text{P}}$ in cyan, violet, and magenta, respectively. The full GW spectra for these points are presented in Fig. 9.8 alongside the PLIS curves for ET, LISA, and NG15. In addition, we show constraints from HLVO3 in terms of its PLIS curve. Constraints from the Planck satellite on the tensor-to-scalar ratio are also indicated.¹³

¹³The construction of the Planck bound in this figure is subtle, and requires further explanation. As detailed in Sec. 9.2.1, we do not specify the inflaton potential at CMB scales, V_{CMB} , but leave it as model-dependent above a given threshold field value φ_{thr} (cf. Eq. (9.36)). Per our construction in Eq. (9.40), we consider that Planck bounds only apply to the more flexible model where $V_{\text{CMB}}(\phi)$ is arbitrary above φ_{thr} . However, to compute the spectra shown in Fig. 9.8 to low frequencies requires a definite potential shape down to CMB scales. The GW spectra we compute therefore assume the definite shape $V(\phi) = m^2 \phi^2 / 2$ even up to CMB scales. We can call this the $m^2 \phi^2$ -equivalent of the actual, more flexible model that we otherwise assume in our discussion. The CMB bound in Fig. 9.8 should then be interpreted as follows: if an $m^2 \phi^2$ -equivalent GW spectrum violates the CMB bound in Fig. 9.8, the corresponding more flexible potential in Eq. (9.36) also violates the observational CMB constraints on A_s , r , and $|\xi_{\text{CMB}}|$, with the converse also being true. This construction firstly entails that the CMB limit in Fig. 9.8 is not a model-independent sensitivity curve on the same footing as the PLIS curves in the plot; it is constructed assuming the $m^2 \phi^2$ -equivalent model, and only

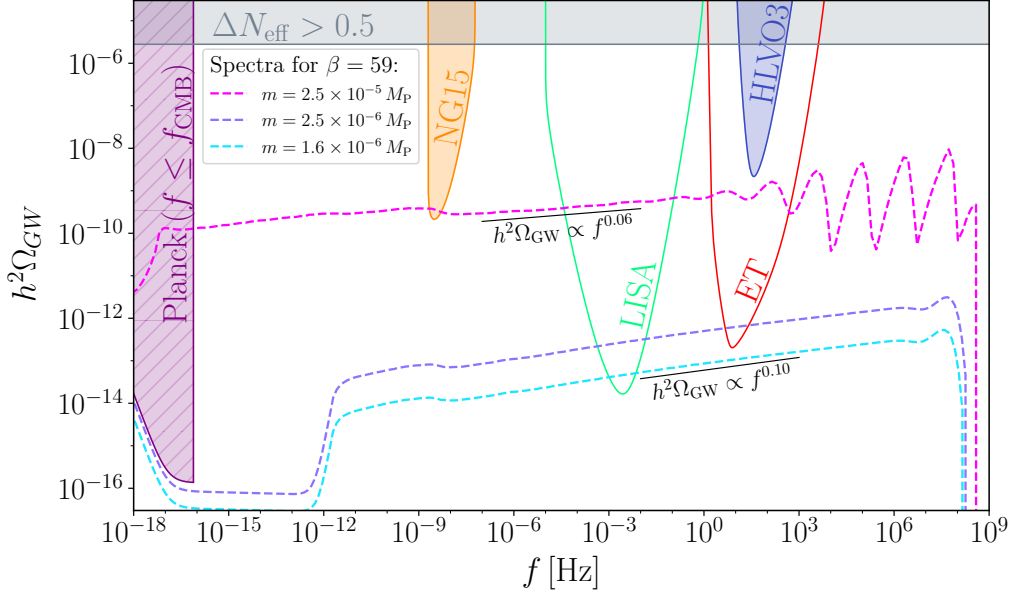


FIGURE 9.8: GW spectra for the benchmark points $\beta = 59$ with $m = 1.5 \times 10^{-6} M_P$ (cyan), $m = 2.5 \times 10^{-6} M_P$ (violet), and $m = 2.5 \times 10^{-5} M_P$ (magenta). PLIS curves are shown for ET, LISA, and NG15 with the same colors as in Fig. 9.7. In addition, we indicate constraints from HLVO3 in blue and from Planck in purple (cf. footnote 13). Clearly, all spectra avoid ΔN_{eff} constraints, but the magenta spectrum conflicts with the Planck observations.

Comparing Fig. 9.8 to its counterpart for PAI, Fig. 9.4, confirms our previous description of the differences between PAI and FAI. For PAI, we found peaked spectra with a strongly blue-tilted GW spectrum, surpassing the vacuum contribution by many orders of magnitude. For FAI, on the other hand, GFIGWs are abundant enough to be observable by ET, LISA, or even NG15, but never in threat of violating bounds on ΔN_{eff} . Instead of the drastic amplification of GWs due to strong backreaction, a temperate amplitude in GFIGWs is sustained over many orders of magnitude in frequency. A first initial burst in gauge-field production lifts the spectrum above the vacuum contribution, where it then remains with only a mild blue tilt. This is clearly a consequence of the dampened efficiency of gauge-field production due to Schwinger pair creation.

The two lower-mass spectra, $m = 1.5 \times 10^{-6} M_P$ and $2.5 \times 10^{-6} M_P$, are nearly identical up to an increase in amplitude for the higher-mass spectrum. For $f \gtrsim 10^{-11}$ Hz, the spectrum rises with a mild blue tilt, $h^2 \Omega_{\text{GW}} \propto f^{0.1}$ (except for a small decrease due to the change in $g_{*,S}$ around the QCD crossover at $f \sim 10^{-9}$ Hz). The spectra reach the sensitivity curves for ET and LISA, and also avoid CMB constraints at $f \sim f_{\text{CMB}}$. The GFIGW contribution only becomes important at $f \sim 10^{-12}$ Hz and the vacuum contribution is not in tension with the upper limit on the tensor-to-scalar ratio, $r \lesssim 0.03$, as indicated by the purple shaded region in Fig. 9.7.

The spectrum at high mass, $m = 2.5 \times 10^{-5} M_P$, is qualitatively different. It exhibits features that we can clearly attribute to fermion-tempered backreaction. As a reminder, fermion-tempered backreaction features a drastically increased duration of inflation, including multiple subsequent oscillation periods in the electromagnetic energy densities. As a reference, the reader is invited to

applies there. Secondly, the $m^2 \phi^2$ -equivalent GW spectrum does not need to satisfy the constraint on the tensor power spectrum, $r A_S \lesssim 0.03 \times 2.1 \times 10^{-9}$, which only applies to the GW spectra computed for the actual, more flexible model, where $V_{\text{CMB}}(\phi)$ is unspecified. What we show in Fig. 9.8 is a “would-be” upper limit on $r A_S$ which only applies to the $m^2 \phi^2$ -equivalent model. To construct it we take Eq. (9.38) and Eq. (9.40) for $f_{\text{thr}} = 10^{-12}$ Hz and $N_{\text{BR}} = 0$. The excluded region is then constructed assuming a flat tensor power spectrum up to $f_{\text{CMB}} = 7.7 \times 10^{-17}$ Hz.

consider Fig. 6.6, which shows the inflationary dynamics for the spectrum we are discussing here. As we can infer from this figure, fermion-tempered backreaction causes a drastically increased duration of inflation for this benchmark point, $\Delta N_{\text{BR}} \simeq 14$. This increased duration of inflation implies that the GW spectrum in Fig. 9.8 is redshifted by an additional factor of 10^5 with respect to the other two spectra in this plot. This explains why this spectrum rises at much lower frequencies compared to the other two. The oscillatory features in electromagnetic radiation are also clearly visible as the repeated peaks at high frequencies.

Notably, there is a clear conflict between the spectrum for $m = 2.5 \times 10^{-5} M_{\text{P}}$ and Planck constraints. This conflict is two-fold. On the one hand, the GFIGW spectrum already rises at CMB scales for this spectrum, thus violating bounds on the tensor-to-scalar ratio r . However, this is model-dependent and is due to our simplified assumption that the inflaton potential is purely $m^2 \phi^2$ -like for this data point. If we were to flatten the inflaton potential at early times, one can expect gauge-field production to set in later due to the lower inflaton velocity on the flatter sections, thus shifting the GFIGW part of the spectrum further to the right. However, this would not save the GW spectrum presented here from being ruled out, as even the vacuum contribution for this spectrum is too high, which can be both read off from Fig. 9.7 and Fig. 9.8 (see also footnote 13).

Another thing to remark upon about the spectrum for $m = 2.5 \times 10^{-5} M_{\text{P}}$ is that it reaches the sensitivity for NANOGrav. However, even if we disregard the fact that this data point is ruled out by Planck data, the spectrum we observe is still not suitable to explain existing NG15 data, despite the fact that it reaches a sufficient amplitude in Ω_{GW} . NG15 data favors a GW signal with a spectral tilt $h^2 \Omega_{\text{GW}} \propto f^{5-\gamma}$, with $\gamma = 3.2 \pm 0.6$ (median and 90% credible interval) [242]. Clearly, the spectrum we see in Fig. 9.8 falls short of this spectral slope with $h^2 \Omega_{\text{GW}} \propto f^{0.06}$ around the relevant frequencies. In fact, changes in the effective number of relativistic degrees of freedom around the QCD phase transition even lead to momentary drops of the spectral index below zero (i.e., $\gamma > 5$), which is clearly disfavored by the NG15 data.

In summary, our results for FAI are in stark qualitative contrast to our results for PAI. For PAI, we found that the most stringent constraints on the (m, β) parameter space came from the bound $\Delta N_{\text{eff}} < 0.5$, which is entirely irrelevant for the FAI model. The reason is that the reduced efficiency of gauge-field production in FAI prevents the drastic overproduction of GFIGWs that we observe for PAI. These more temperate dynamics result in GW spectra which are prospectively observable by both ET and LISA. However, these same damping effects also imply that a larger Hubble rate at production is required in order to supply the necessary amplitude in GWs. Therefore, we find that the most relevant constraints on the SGWB from FAI come from the Planck measurements of A_S and the limit on r . Similarly, these damping effects entail that it is hard to achieve a sufficiently large amplitude in Ω_{GW} to reach the sensitivity of NG15. To achieve this amplitude, one needs to enhance the GW amplitude by increasing the inflaton mass, which conflicts with Planck data, and, regardless of these bounds, also fails to achieve a spectrum which is sufficiently steep to agree with NG15 results.

9.4.3 Estimating fermion-induced gravitational waves

Up until now we have exclusively considered GWs sourced by gauge fields. However, our previous exploration of the FAI model in chapter 6 has served to demonstrate that the Schwinger effect can lead to the production of a large abundance of fermions during axion inflation, with their energy density possibly surpassing that of the gauge field at the end of inflation. If this fermion gas is sufficiently anisotropic, one would therefore expect that it contributes correspondingly to the total amplitude of the GW spectrum.

A computation of this additional contribution, on equal footing to the one for the gauge field, is beyond the scope of this thesis. But, we can make an educated estimate of its relevance based on a

simple scaling argument that accounts for the amplitude of the gauge-field-induced GW spectrum, and the energy density that is, respectively, contained in fermions and gauge fields:

$$\Omega_{\text{GW}}^{\chi} \sim \alpha_* \left(\frac{\rho_{\chi}}{\rho_{\text{EM}}} \right)^2 \Omega_{\text{GW}}^{\text{EM}}. \quad (9.48)$$

Here, $\Omega_{\text{GW}}^{\chi}$ is the contribution from fermions to the GW spectrum, while $\Omega_{\text{GW}}^{\text{EM}}$ is that of the gauge field. What we assume here is that the overall spectral dependence for the fermion contribution matches that of the gauge field, and we simply relate the two contributions via their respective energy densities, ρ_{χ} for the fermions, and ρ_{EM} for the gauge field. This assumption aligns with our effective modelling of the conducting fermionic medium as direct functions of the gauge-field expectation values, $\langle \mathbf{E}^2 \rangle$, $\langle \mathbf{B}^2 \rangle$, and $\langle \mathbf{E} \cdot \mathbf{B} \rangle$ (cf. chapter 6). The additional coefficient α_* is included to describe how inherently anisotropic (quadrupolar and beyond) the fermion gas is, i.e., its ability to source GWs.¹⁴

Our GEF results give us direct access to the evolution of ρ_{EM} and ρ_{χ} , and also allow us to access $\Omega_{\text{GW}}^{\text{EM}}$. What is missing to evaluate Eq. (9.48) is an estimate for α_* , which requires knowledge of the microphysical behavior of fermions. To estimate α_* , we should reflect on the origin of fermion production in our model. As discussed in chapter 6, fermion production in axion inflation proceeds via the Schwinger mechanism for nearly collinear electric and magnetic fields. As argued in Ref. [120], this pair creation process results in the production of two distinct fermion populations. The first population are chiral fermions which are directly produced from the electric field via the chiral anomaly. Ref. [120] argues that these fermions quickly thermalize due to internal scattering, which we take to imply that they quickly become isotropic and are thus irrelevant for the production of GWs. However, a second population of left-right symmetric fermions occupies higher Landau levels due to the presence of the magnetic field. For these fermions, scattering is negligible, and they do not thermalize. In Ref. [120], it is estimated that this second population is dominant in ρ_{χ} , with chiral fermions only making up a small fraction of around $10^{-2} - 10^{-3}$. We take this to imply that, in principle, a large fraction of ρ_{χ} may be anisotropic. We will thus assume an optimistic value, $\alpha_* = 1$ for our estimates.

In Fig. 9.9, we show the prospective relevance of the fermionic contribution by considering how the SNR regions for LISA (green), ET (red), and NG15 (orange) are affected when adding the estimate for $\Omega_{\text{GW}}^{\chi}$ from Eq. (9.48) to $\Omega_{\text{GW}}^{\text{EM}}$ from Eq. (9.25). The $S/N > 1$ regions enveloped by dotted lines indicate the result with $\Omega_{\text{GW}}^{\chi}$ included, while the dashed-dotted lines are computed from $\Omega_{\text{GW}}^{\text{EM}}$ alone. Besides LISA, ET, and NG15, also HLVO3 now shows sensitivity to Ω_{GW} . The sensitivity region for HLVO3 is shown in blue.

There are several notable effects of the fermionic contribution to Ω_{GW} . Firstly, we may observe that fermion-sourced GWs increase the sensitivity region for LISA and ET, particularly for large β . This clearly signals that, in the frequency regions covered by LISA and ET, $\Omega_{\text{GW}}^{\chi}$ may contribute more to the total spectrum than $\Omega_{\text{GW}}^{\text{EM}}$. That this effect is more pronounced for larger β may be readily attributed to the fact that an increase in β increases the efficiency of gauge-field production via the inflaton field, which in turn increases fermion production, such that $\Omega_{\text{GW}}^{\chi}$ becomes more relevant as well. In contrast to ET and LISA, we find that the sensitivity region for NG15 has not been significantly increased. We attribute this to the observation that fermion production occurs towards the end of inflation, as suggested by our results in Sec. 6.4. Therefore, at low frequencies, $\Omega_{\text{GW}}^{\chi}$ contributes less to the total GW spectrum than $\Omega_{\text{GW}}^{\text{EM}}$ does, so the Ω_{GW} in the nanohertz

¹⁴The inspiration for this Ansatz comes from Ref. [470], who derived a heuristic formula for estimating the GW amplitude for an arbitrary GW source after inflation. Their formula does not directly apply to our case, because we consider inflationary GW production, but their scaling arguments should apply. These entail the quadratic scaling of $\Omega_{\text{GW}}^{\chi}$ with energy density, and the inclusion of the parameter α_* to parametrize the anisotropy of the source.

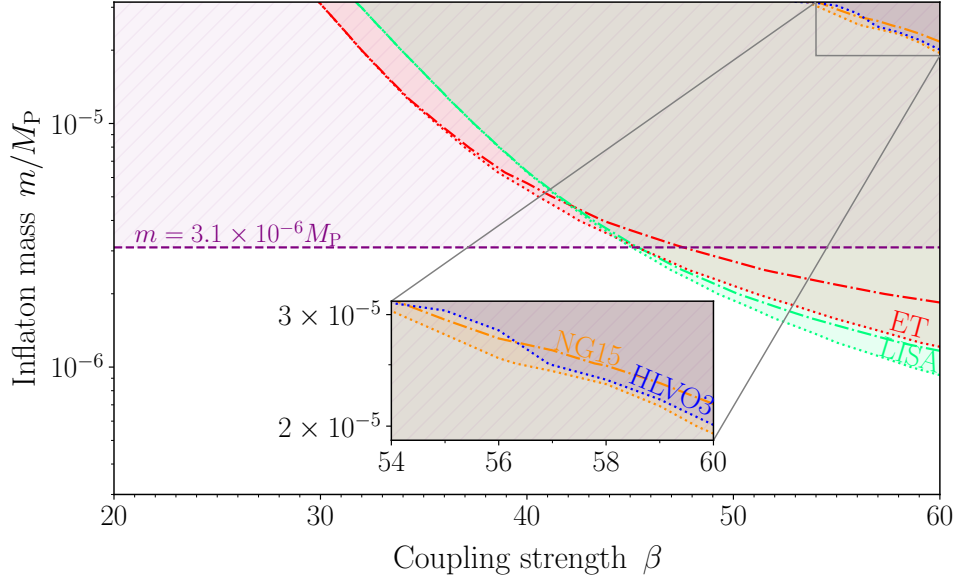


FIGURE 9.9: Sensitivity regions when estimating the additional contribution to $h^2\Omega_{\text{GW}}$ due to the fermion gas. The regions with $S/N > 1$ are shown for ET (red), LISA (green), NG15 (orange), and HLVO3 (blue). The fermion contribution may boost the detectability region for ET and LISA, thus dominating over the GFIGW contribution at high frequencies. At low frequencies, the extra contribution matters little, as evident by the NG15 region being nearly insensitive to the additional source. More so, the enhanced signal at high frequencies implies that HLVO3 is in conflict with a FAI interpretation of the NG15 signal. These results are a simple estimate assuming an optimistically large fraction of anisotropic fermions.

band remains largely unaffected. However, at late times, the fermionic contribution becomes more important, and thus $\Omega_{\text{GW}}^{\chi}$ can surpass $\Omega_{\text{GW}}^{\text{EM}}$ at higher frequencies. This explains the appearance of an $S/N > 1$ region for HLVO3: because $\Omega_{\text{GW}}^{\chi}$ is more blue-tilted than $\Omega_{\text{GW}}^{\text{EM}}$, the total amplitude in GWs is overall increased at higher frequencies. Based on our simple estimates alone, including $\Omega_{\text{GW}}^{\chi}$ thus further weakens the FAI interpretation of the NG15 signal, as the HLVO3 region largely overlaps with the one for NG15.

9.5 Summary and conclusion

In this chapter, we have analyzed the creation of gravitational waves in axion inflation with a focus on the dynamical balance between gauge-field creation, backreaction, and Schwinger pair creation. Using the gradient-expansion formalism (GEF), we investigated gauge-field-induced gravitational waves (GFIGWs) in two models of axion inflation; pure axion inflation (PAI), which does not account for Schwinger pair creation, and fermionic axion inflation (FAI), which does. Based on this framework, we then assessed how the different dynamical mechanisms in these models impact the resulting gravitational-wave (GW) signal by performing a comprehensive phenomenological analysis of its detectability for next-generation GW interferometers and PTA experiments.

Our results for PAI demonstrate the importance of gauge-field backreaction effects in this model. Studying GFIGW production in the parametric regime close to the onset of strong backreaction, we find that an observable GW signal in the Einstein Telescope (ET) or the Laser interferometer Space Antenna (LISA) is only realized in regions of parameter space that also lead to strong backreaction. At the same time, we observe that the non-linear dynamics of strong backreaction

result in an extended period of gauge-field production during which the electromagnetic background is considerably amplified and the resulting induced GWs end up conflicting with BBN and CMB bounds on the number of additional relativistic species in the early Universe, ΔN_{eff} .

Our results for FAI instead underpin the importance of accounting for the nontrivial interplay between gauge-field production and the Schwinger effect during axion inflation. In the particular case we have considered here, i.e., one where the Schwinger effect is highly effective, we find that Schwinger pair creation demonstratively drains the electromagnetic background, which therefore inhibits the ability of GFIGW production. This effect however does not imply that GFIGWs from FAI are unobservable; our results demonstrate that GW production during FAI can source a stochastic gravitational wave background (SGWB) in reach of ET or LISA as long as inflation occurs at sufficiently high scale. We even find GW signals with a sufficient amplitude to reach the sensitivity range of the latest dataset of the NANOGrav collaboration (NG15). However, this prediction necessitates an inflationary scale in tension with existing Planck data, and the resulting spectral slope is too shallow to adequately explain NG15 observations. Based on simple estimates, we also note that GWs sourced by fermions may be highly relevant in this model, as they could even dominate the GW spectrum at high frequencies. This could further increase the prospects of detecting an SGWB from FAI with LISA or ET. Moreover, this could entail additional pressure on an interpretation of the NG15 dataset in terms of GWs from FAI, seeing as this signal is correlated with a high-frequency fermionic contribution to the GW spectrum in conflict with the non-observation of a SGWB by LIGO and Virgo. In stark contrast to our findings for PAI, we observe that strong-backreaction effects are not required for a GW signal from FAI to reach the sensitivity of LISA or ET. Furthermore, we find that the damped production of gauge fields in FAI results in mildly blue-tilted GW spectra which are never in threat of violating ΔN_{eff} bounds.

For our analysis, we relied on simplifying assumptions of a homogeneous inflaton field even when modelling the regime of strong backreaction. As established in previous chapters of this thesis, this assumption is necessary for us to solve the dynamics of axion inflation using the GEF, but it also entails that our analysis does not capture the dynamical growth and backreaction of inflaton inhomogeneities in the regime of strong gauge-field backreaction. This is particularly relevant for our analysis of PAI, where lattice simulations demonstrate that inflaton gradients are dynamically relevant during strong backreaction. This needs to be taken into consideration when interpreting our results for this model. For one, our observation that GWs from PAI are only observable by LISA and ET if strong-backreaction effects are present appears to be solid, seeing as GFIGWs produced during weak backreaction fall short of the required amplitude, and the GEF method is well-suited to study these dynamics. What is less certain is that ΔN_{eff} violation is indeed an inevitability of the strong-backreaction period, seeing as inhomogeneous-backreaction effects may prevent an overproduction of GFIGWs. Indeed, based on the recent lattice results of Ref. [380], it seems likely that the additional impact of inflaton gradients allows for a smoother transition between the weak and strong-backreaction regime than what our results suggest. Such a more gradual rate of gauge-field production in the strong-backreaction regime may lead to a detectable GW signal for PAI in ET, LISA, or even NANOGrav after all. In this case, the results we have presented here may serve as a guide for identifying the relevant regions in parameter space for further studies using, e.g., lattice simulations.

For the model of FAI, the role played by inflaton inhomogeneities is less certain given the lack of lattice studies for this model. Seeing as our results are not as dependent on strong-backreaction effects, it is however conceivable that these effects may be less impactful than for PAI. Also in the case of FAI, our findings beckon the need for further scrutiny using lattice techniques. Similarly, our effective modelling approach to fermion creation deserves further attention. Our results emphasize the potential relevance of fermion production during axion inflation, which inadvertently reinforces the necessity to go beyond our effective approach and to compute fermion backreaction onto the

gauge field from first principles.

A natural continuation of our work is to extend our analysis to other variants of axion inflation. In our work, we assumed that the inflaton potential is quadratic near its minimum, but our conclusions may be altered if the inflaton potential is not parabolic at times when tensor modes relevant for GW observations exit the Hubble horizon, see for instance Refs. [82–84]. Moreover, applying our formalism to multifield realizations of axion inflation could yield potentially interesting results, for instance for the case of aligned natural inflation [88], or even in the scenario where the axion field which sources gauge fields is a spectator to inflation [84, 471, 472].

Another topic of interest concerns GFIGW production during the reheating phase, which was previously studied by Refs. [85, 380, 385, 386]. Our results motivate the further investigation of these effects for PAI also in previously unexplored regions of parameter space indicated by our GEF benchmark. More so, preheating studies for the model of FAI are still missing at the current point in the literature, and may reveal important additional effects for GW production, and beyond.

Our focus in this chapter has primarily been on the gauge field as a source for an SGWB. For the model of FAI, we have already pointed out that the fermions themselves could make a relevant contribution to the SGWB. The other major source of GWs which could meaningfully contribute to the spectrum of GWs from axion inflation are scalar perturbations, which can source GWs at second order in perturbation theory. In the model of PAI, these scalar perturbations can be efficiently sourced by gauge fields, while the fermions in FAI may also act as a source of scalar perturbations. Additional features in the GW spectrum due to scalar-induced or fermion-induced gravitational waves could alter the detection prospects for GWs from axion inflation beyond what we have reported here. A starting point for investigating scalar-induced gravitational waves is necessarily the study of scalar perturbations from gauge fields, which entails building on our results in the previous chapter. For the case of FAI, our estimates of the fermion-induced contribution to the GW spectrum clearly motivates the further study of both fermion-induced scalar and tensor perturbations. A starting point in this direction could be to apply the formalism for computing fermion-induced GWs presented in Refs. [122, 123] to the model of FAI.

Overall, the results of our analysis demonstrate how the rich dynamical structure of axion inflation, both with and without the inclusion of the Schwinger effect, can profoundly affect phenomenological prediction for axion inflation like the sourcing of gravitational radiation. This underpins the necessity for apt numerical techniques to resolve these dynamics, and further reinforces our development of the GEF formalism both for pure axion inflation and fermionic axion inflation in previous chapters of this thesis. Although our analysis of pure axion inflation substantiates the need for capturing the full dynamics of the strong-backreaction regime, including the influence of inflaton gradients, it also highlights the relevance of homogeneous-backreaction techniques like the GEF; with its relatively cheap numerical cost and large dynamical range, it allows one to perform a comprehensive analysis of large regions of our model’s parameter space, which can be used to guide the numerically more costly lattice simulations in follow-up studies.

In the case of FAI, our effective treatment of Schwinger pair creation with the GEF allowed us to reveal that, while Schwinger damping drastically reduces gauge-field production, it does not render the effects of gauge-field production phenomenologically unimportant. In fact, our results suggest that the opposite is true: the damped creation of gauge fields allows one to avoid undesirable consequences such as the overproduction of gravitational radiation, while the additional creation of fermions may add even more wealth to the study of GWs from axion inflation than previously expected. These results motivate a rich research program to refine the modelling of inflationary dynamics including the creation of gauge fields and the Schwinger effect, and to further investigate the direct phenomenological implications of fermion production.

Conclusion

Cosmic inflation is the leading paradigm in modern cosmology to explain the origin of primordial inhomogeneities whose existence is evidenced in the anisotropies of the cosmic microwave background (CMB). In the conventional models of slow-roll inflation, this process is modelled as a minimally coupled scalar field, which drives a period of accelerated expansion, while its vacuum fluctuations, and those of the spacetime metric, generate the curvature perturbations that seed the CMB anisotropies. Similarly, excited vacuum fluctuations of the metric also give rise to primordial gravitational waves (GWs), which may be detected by GW experiments, or future CMB missions. However, in realistic extensions of slow-roll inflation, the inflaton field can also couple to other matter fields, which may modify these conventional predictions. Such interactions can trigger the creation of particles during inflation, which can in turn source primordial perturbations in addition to the conventional perturbations that arise from vacuum fluctuations. Understanding the impact that inflationary particle production can have on observables is an important theoretical undertaking, as distinctive features of particle production in, e.g., the spectra of inflationary GWs, may be experimentally probed in current or next-generation GW experiments.

In this context, models of axion inflation occupy an important role, because the pseudoscalar inflaton field in these models may couple to Abelian gauge bosons via the Chern–Simons operator $\phi F_{\mu\nu} \tilde{F}^{\mu\nu}$. This is an effective channel for Abelian gauge-field production, through which the rolling of the inflaton field can effect a near exponential production of helical gauge bosons. These field excitations may also be a channel for continued particle production, as they can create particle–anti-particle pairs from the vacuum via the Schwinger effect. At the same time, gauge-field production in axion inflation is also prone to significant backreaction on the inflaton dynamics, while charge carriers may backreact on gauge-field production. A self-consistent treatment of these coupled dynamics requires a numerical framework capable of evolving the inflationary background and capturing the interplay between these interacting sectors, while also connecting the resulting dynamics to observables such as the spectra of primordial curvature perturbations and GWs.

In this thesis, we have developed such a self-consistent numerical framework by building on the *gradient-expansion formalism* (GEF) [128, 135]; a semi-analytical formalism to describe gauge-field production and backreaction during axion inflation. The GEF reformulates the nonlinear equations of motion of axion inflation into a coupled set of ordinary differential equations (ODEs), which are numerically inexpensive to solve. Prior to the work presented in this thesis, the GEF had existed as a complementary method to the existing numerical methods for studying axion inflation, i.e., the mode-by-mode approach and lattice simulations of axion inflation, but it had only been applied in select case studies that mainly demonstrated its ability to solve the equations of motion of axion inflation [128]. Throughout this thesis, we have distinctly changed this status quo: as part of the four analyses we have presented, the GEF has developed into a tool that can be used for comprehensive stability analysis in axion inflation (cf. chapter 5), a tool which models the Schwinger effect in a more refined way (cf. chapter 6), and, we have expanded the GEF into a framework that can be directly applied to infer both the spectrum of gauge-field-induced scalar (cf. chapter 8) and tensor (cf. chapter 9) perturbations based on the background dynamics that are determined through the GEF. These developments have all been performed as part of specific research projects whose individual results are important contributions to the literature on

axion inflation in their own right. For the conclusion of this thesis, we now want to view these contributions through the lens of the GEF as a self-consistent numerical *GEF framework* that can be applied to the study of axion inflation.

Taking this perspective, the improved effective model of Schwinger pair production that we have devised in chapter 6 is a cornerstone in the development of this numerical GEF framework. Our revised effective description builds upon the existing modelling of Schwinger pair production in axion inflation, which assumes that the charge carriers produced by the Schwinger effect constitute a conductive medium that dampens gauge-field production. Our improved version addresses two conceptual shortcomings of this original description. First, we resolved an ambiguity in the effective description of charge carriers as a conductive medium. Previous authors modelled the backreaction of charge carriers onto gauge-field production as either a charge current expressed through a purely electric conductivity [104, 128], $\mathbf{J} = \sigma_E \mathbf{E}$, or a purely magnetic conductivity [120, 121, 132], $\mathbf{J} = \sigma_B \mathbf{B}$. This ambiguity arises from the idealization that electric and magnetic fields in axion inflation are collinear. We improved upon this by deriving an unambiguous mixed representation, $\mathbf{J} = \sigma_E \mathbf{E} + \sigma_B \mathbf{B}$, which is valid for the realistic case where electric and magnetic fields are not collinear in the comoving frame. In addition, we also identified a characteristic length-scale l_S that describes the separation of charge carriers in the conductive medium given the strength of the electric field that has created them. Prior to our work [128], Schwinger damping was modelled as affecting gauge-field modes of arbitrary wavelength, λ . We modify this description such that only gauge-field modes with wavelengths $\lambda > l_S$ are subjected to the presence of this conductive medium. This reflects a more realistic physical description of Schwinger damping.

Collectively, these two modifications; scale-dependent damping, and the mixed-picture representation of the induced current $\mathbf{J}$, represent a conceptual improvement over the existing effective description of Schwinger damping. By adapting the GEF to incorporate these changes, we could therefore improve our numerical GEF framework in its ability to dynamically determine the effect of Schwinger pair production. In a first application, this revised version of the GEF enabled us to perform a comprehensive case study regarding the efficiency of Schwinger pair production in the case where the gauge field in axion inflation produces massless Standard Model (SM) fermions. As part of this case study, we could demonstrate that our new effective model of Schwinger damping can lead to order-of-magnitude differences in the energy density stored in both gauge fields and fermions. Moreover, the GEF also enabled us to identify a previously unknown dynamical regime of *fermion-tempered backreaction*, where efficient Schwinger pair production coexists with strong gauge-field friction effects on the inflaton field.

A second key component of our work is the development of a numerical scheme that can be used to determine scalar and tensor power spectra for axion inflation based on the inflationary background dynamics that we computed using the GEF. In particular, our method permits us to determine both vacuum and gauge-field-induced contributions to the scalar and tensor power spectra, which had previously not been done using the GEF.

As a first application, this permitted us to closely assess the role of gauge-field-induced metric perturbations in axion inflation. Expanding upon the results of previous calculations, we devised a formalism that self-consistently takes into account curvature perturbations coming from gauge-field sourced inflaton *and* scalar metric perturbations, whereas previous such calculations only included sourced inflaton perturbations. Building on the results of a controlled analytical calculation, the numerical approach based on the GEF allowed us to demonstrate that gauge-field-sourced metric perturbations contribute only marginally to the total spectrum of curvature perturbations in axion inflation. In particular, to reach this conclusion, our numerical GEF framework was vital, as it allowed us to also extend our analysis to a regime where gauge-field backreaction effects on the background dynamics of axion inflation are strong. Based purely on analytical techniques, this

conclusion would otherwise have only been valid assuming negligible gauge-field backreaction on the dynamics of the inflaton field. Similarly, the GEF also permitted us to extend the application of our new formalism to determine how gauge-field backreaction affects the evolution of curvature perturbations themselves. Here, we could show that the inclusion of these effects can lead to unanticipated infrared divergences in the scalar power spectrum. While the significance of this infrared divergence remains unclear, our GEF framework will certainly be a valuable tool in understanding it in the future.

Our work in chapter 9 also serves to illustrate the use of our numerical GEF framework in directly connecting the dynamics of axion inflation to observable predictions. Therein, our framework enabled us to perform a comprehensive study regarding the detectability of GWs from axion inflation with current or next-generation GW observatories. In particular, the numerical efficiency of the GEF allowed for an extensive parameter-space exploration to study observational prospects of gauge-field-induced gravitational waves (GFIGWs) with either current pulsar-timing-array (PTA) experiments or next-generation GW interferometers, all while taking into consideration constraints on inflation from the *Planck* satellite, and bounds on the number of dark relativistic species beyond SM radiation, ΔN_{eff} .

For this study, the analysis pipeline we have devised as part of our GEF framework allowed us to perform an equally thorough investigations into GW production for two separate variations of axion inflation: pure axion inflation (PAI), where the Schwinger effect is treated as ineffective or absent, and fermionic axion inflation (FAI), where we assume that the Schwinger effect produces massless SM fermions. In particular, this comprehensive analysis of GW production in both PAI and FAI serves to underpin the qualitatively distinct phenomenology that may arise when the Schwinger effect is active during axion inflation.

In the analysis of GWs from PAI, we find that a sufficient amplitude in gauge-field-induced gravitational waves (GFIGWs) is only reached once backreaction effects have become important. However, simultaneously, these backreaction effects increase the duration of inflation, which allows for an increased production of GFIGWs. This continued phase of gauge-field production ultimately results in an overproduction of these GWs in violation of existing ΔN_{eff} -bounds on their integrated amplitude. This leads us to conclude that detection prospects for GWs from PAI for both the Laser Interferometer Space Antenna (LISA) and the Einstein Telescope (ET) may be limited.

In contrast, our results indicate a more optimistic perspective on the detectability of GFIGWs from FAI. In this case, we could demonstrate that a damping of gauge-field production effectively reduces the amplitude of the GW spectrum sourced by these fields. Consequently, we find no conflict with GFIGW production and ΔN_{eff} bounds for FAI. Meanwhile, we could show that the amplitude of GWs in this model may be detectable in the LISA and ET frequency band if the scale of inflation is sufficiently high. Moreover, based on simple estimates, we find that these detection prospects could even be enhanced when accounting for fermions as an additional source of GWs at high frequencies. However, such estimates will require a more refined computation in the future, e.g., by directly computing the GW contribution from fermions on a first-principles basis as suggested in Refs. [122, 123].

To our knowledge, the results of the FAI analysis are the first-ever numerical analysis of GWs from axion inflation where Schwinger damping is fully taken into account. Moreover, the analysis of PAI is the first-ever comprehensive parameter space exploration of this kind. Therefore, both these results constitute important contributions to the current literature regarding GW production from axion inflation in their own right, and the GEF framework we have developed in this thesis was vital component in obtaining these results.

Although these different applications of our GEF framework serve to demonstrate its usefulness, our approach does not come without limitations. Fundamentally, all our results and conclusions are

computed based on the inflationary background dynamics that we determine using the GEF. However, at present, the GEF is unable to resolve the dynamical effects that inflaton inhomogeneities can have on the production of gauge fields, and ultimately on the entire inflationary system. As demonstrated in several independent lattice simulations [378–380], this inhomogeneous backreaction becomes dynamically relevant once gauge-field backreaction effects are strong. This can lead to quantitative differences regarding the dynamical evolution of gauge fields during axion inflation as discussed in chapter 4. Since the GEF cannot capture these effects, conclusions that we can draw based on our GEF framework may require further scrutiny in the future using lattice techniques.

This limitation is especially relevant for the case of PAI, where the effect of inhomogeneous backreaction are clearly demonstrated. In this case, the results that one obtains based on our GEF framework should be rather viewed as establishing a *GEF benchmark*, with results that are obtained in the strong-backreaction regime requiring an independent confirmation using lattice simulations. In the context of this thesis, this is most clearly the case for our results regarding the detectability of GWs from PAI, as discussed in detail in chapter 9. However, we believe that these GEF benchmarks offer valuable information for lattice studies, as these simulations can be numerically costly and are thus less suitable for vast parameter-space scans like the ones we can perform with the comparatively faster GEF.

From this perspective, the linearized gradient expansion formalism (LGEF) that we developed and applied in chapter 5 could prove to be an insightful tool. The LGEF can be used to study the linear stability analysis of steady-state solutions in axion inflation, which exist under simplifying assumptions of a constant Hubble rate and an inflaton potential with a constant gradient. One important insight that the LGEF provides is a new criterion to determine the onset of strong-backreaction effects. In the future, this new criterion could be used alongside the GEF to signal the onset of backreaction effects, and to highlight at what point results that are obtained using our GEF framework need to be reassessed by lattice simulations.

An additional insight that the LGEF provides is that it clarifies the stability of the Anber–Sorbo solution — the particular steady state-solution that exists in the regime where gauge-field friction on the inflaton field dominates over Hubble friction. In this context, the LGEF allowed us to refine the previous findings of Ref. [131] that imply that the Anber–Sorbo solution is unstable. Our results give a more subtle perspective: for one, we find that the Anber–Sorbo solution is indeed unstable in the regions of the axion-inflation model space that were explored in Ref [131], but on the other hand, the LGEF reveals a possible region of the axion-inflation phase space, where backreaction may be strong. Enticingly, this region may also correspond to a regime where axion gradients are ineffective at backreacting on the dynamics of gauge-field production. This claim needs to be further analyzed using lattice techniques, but, if true, such regions of stable backreaction are precisely where the GEF could be used to its full capability.

An important open question that should be further explored in future work is the relevance of inhomogeneous-backreaction effects in the context of Schwinger damping. At present, lattice simulations that take the Schwinger effect into consideration appear to indicate that dynamical feedback from inflaton inhomogeneities on gauge-field production is unimportant as long as the Schwinger pair creation is effective at suppressing gauge-field backreaction on the inflaton field [134]. However, this claim needs to be further explored, in particular, since we have demonstrated in this work that Schwinger damping and gauge-field backreaction can coexist in the case of fermion-tempered backreaction. The relevance of inhomogeneous-backreaction effects in this dynamical regime requires careful scrutiny.

One possibility to approach this problem could be a comparative analysis between the GEF and lattice simulations including the Schwinger effect, similar to the thorough comparison for PAI performed in Refs. [378, 379]. In this case, the GEF may prove useful in calibrating such

simulations. As we have seen explicitly in our work, Schwinger damping accumulates through the course of inflation over a large range of e -folds, which could be challenging to resolve in lattice simulations due to their limited dynamical range. Here, the GEF could be used to calibrate these simulations, as it can provide the necessary nontrivial input regarding the dynamical state of the coupled system consisting of the inflaton field, the gauge field, and charged fields.

An alternative possibility to assess the dynamical relevance of inhomogeneous backreaction would be to compute their effect perturbatively. That is, one may first compute the dynamics of inflation based on the GEF, assuming inflaton inhomogeneities are dynamically irrelevant. In a second step, one then determines the amplitude of inflaton inhomogeneities to assess their potential to backreact on gauge-field production. A similar approach was recently taken in Ref. [87] who studied GW production in PAI using the mode-by-mode approach. Withing our GEF framework, such an *a posteriori* assessment regarding the importance of inflaton inhomogeneities could be achieved by building on our existing method to compute the spectra of sourced scalar perturbations. Indeed, this approach could also be used to assess the relevance of inflaton inhomogeneities in the presence of Schwinger damping even without relying on lattice simulations.

Moreover, a numerical approach that can perturbatively assess the importance of inhomogeneous backreaction could also prove valuable in combination with the LGEF, for instance to confirm or refute the claim that the LGEF can be used as an indicator of inhomogeneous backreaction, as we have speculated upon in chapter 5.

Evidently, inhomogeneous backreaction is a highly relevant topic when discussing numerical methods that aim at studying the dynamical production and backreaction of gauge fields during axion inflation. Ultimately, it may therefore be worthwhile to revisit the results of Ref. [372], who attempt to perturbatively include inflaton-backreaction effects into the GEF. As we have noted upon in chapter 4, their approach may be refined by taking into account longitudinal gauge-field polarizations in the existing GEF formulation, as inflaton inhomogeneities can also source these modes [379]. Including longitudinal gauge-field modes into the GEF could help to improve the agreement between lattice and GEF results in the strong-backreaction regime, and would ultimately improve the entire GEF framework we have established in this thesis. A starting point for this improvement could be the GEF variant developed in Ref. [392], which accounts for the production of longitudinal gauge-field modes in an inflation scenario featuring massive vector fields.

Beyond the discussion of inhomogeneous-backreaction effects, the numerical GEF framework we have devised in this thesis could also serve as the basis for a variety of phenomenological analyses that expand upon our current work. For instance, our method for computing the spectrum of curvature perturbations using the GEF could be used as a starting point into a thorough analysis of primordial-black-hole formation in axion inflation. Similarly, our approach could also be used to determine the importance of scalar-induced gravitational waves in axion inflation, an effect which our GW analysis did not yet account for. Finally, our improved modelling of the Schwinger effect may serve as the basis for exploring dark-matter production in axion inflation, or to consider SM fermion production in axion inflation in the context of baryogenesis scenarios.

Overall, the concrete applications of our GEF framework in this thesis have established it as a useful tool, which can now be used in a variety of applications beyond the results of this thesis. The essence of this GEF framework is our new Python package, the GEFF [7], wherein many improvements, modifications, and developments that we have devised in this thesis are self-consistently implemented. With this, we hope to provide a basis on which future work can build to further explore the rich phenomenology of axion inflation, and to improve our understanding of its complex dynamical structure.

Appendices

Appendix A

Whittaker functions

In this appendix, we briefly summarize the properties of Whittaker functions to the extent that they are needed for the discussion of axion inflation in chapters 3, 6, and 8. The properties of Whittaker functions may be, e.g., found in Refs. [186, 473].

A.1 Properties of Whittaker functions

The two kinds of Whittaker functions, $M_{\kappa,\mu}$ and $W_{\kappa,\mu}$ constitute a set of linearly-independent solutions to the Whittaker differential equation

$$\frac{d^2 f}{dz^2} + \left(-\frac{1}{4} + \frac{\kappa}{z} + \frac{1/4 - \mu^2}{z^2} \right) f = 0. \quad (\text{A.1})$$

These two solutions are related to the confluent hypergeometric Kummer Φ and Tricomi U functions as

$$M_{\kappa,\mu}(z) = e^{-z/2} z^{\mu+1/2} \Phi(\mu - \kappa + 1/2; 1 + 2\mu; z), \quad (\text{A.2})$$

$$W_{\kappa,\mu}(z) = e^{-z/2} z^{\mu+1/2} U(\mu - \kappa + 1/2; 1 + 2\mu; z). \quad (\text{A.3})$$

In the limit of $|z| \rightarrow \infty$ (with $-3\pi/2 < \arg z \leq -\pi/2$), their asymptotic form is given by

$$\begin{aligned} M_{\kappa,\mu}(z) &= \frac{\Gamma(2\mu + 1)}{\Gamma(\mu + 1/2 + \kappa)} e^{-i\pi(\mu+1/2-\kappa)} e^{-z/2} z^\kappa [1 + O(z^{-1})] \\ &\quad + \frac{\Gamma(2\mu + 1)}{\Gamma(\mu + 1/2 - \kappa)} e^{z/2} z^{-\kappa} [1 + O(z^{-1})], \end{aligned} \quad (\text{A.4})$$

$$W_{\kappa,\mu}(z) = e^{-z/2} z^\kappa [1 + O(z^{-1})]. \quad (\text{A.5})$$

Here, Γ denotes the Euler Gamma function. The derivatives of $M_{\kappa,\mu}$ and $W_{\kappa,\mu}$ further satisfy the recurrence relationships

$$z \frac{d}{dz} M_{\kappa,\mu}(z) = \left(\frac{z}{2} - \kappa \right) M_{\kappa,\mu}(z) + \left(\frac{1}{2} + \mu + \kappa \right) M_{\kappa+1,\mu}(z), \quad (\text{A.6})$$

$$z \frac{d}{dz} W_{\kappa,\mu}(z) = \left(\frac{z}{2} - \kappa \right) W_{\kappa,\mu}(z) - W_{\kappa+1,\mu}(z). \quad (\text{A.7})$$

Finally, the Wronskian of the Whittaker functions takes the form

$$\mathcal{W}[W_{\kappa,\mu}(z), M_{\kappa,\mu}(z)] = \frac{\Gamma(2\mu + 1)}{\Gamma(\mu + 1/2 - \kappa)}, \quad (\text{A.8})$$

as may be readily derived from Eqs. (A.4) and (A.5) and the fact that the Wronskian is a constant function of z .

A.2 Whittaker functions in axion inflation

The properties of Whittaker functions are needed to determine analytical expressions for gauge-field mode functions in axion inflation. For the purposes of this thesis, we require these expressions in two particular scenarios: the conventional axion inflation model without Schwinger pair production, which we discuss in chapter 3, and axion inflation including Schwinger pair creation, discussed in chapter 6.

The expressions we give below may be, for instance, used to determine boundary terms appearing in the gradient-expansion formalism equations in chapters 4 and 6.

Axion inflation without the Schwinger effect

The conventional mode equation for a gauge-field mode $A_\lambda(\eta, k)$ in axion inflation takes the form (c.f. Eq. (3.19))

$$A_\lambda''(\eta, k) + (k^2 - 2\lambda\xi aHk) A_\lambda(\eta, k), \quad (\text{A.9})$$

with η conformal time, k the mode functions wavenumber, λ its polarization, and ξ the instability parameter defined in Eq. (3.20). Assuming de Sitter space, $\eta = -1/(aH)$, and a constant value for ξ , Eq. (A.9) takes the form of the Whittaker equation in Eq. (A.1) (with $z = -k\eta$, $\mu = \pm 1/2$, and $\kappa = -i\lambda\xi$). Imposing Bunch–Davies initial conditions, $A_\lambda(\eta, k) \rightarrow e^{-ik\eta}/\sqrt{2k}$ for $-k\eta \rightarrow \infty$, selects the solution given by $W_{\kappa, \mu}$ as may be determined from its asymptotic form in Eq. (A.5). Hence, one obtains

$$A_\lambda(\eta, k) = \frac{1}{\sqrt{2k}} e^{\frac{\lambda\pi\xi}{2}} W_{-i\lambda\xi, \frac{1}{2}}(2ik\eta), \quad (\text{A.10})$$

in agreement with Eq. (3.28).

Axion inflation including the Schwinger effect

For the effective model of axion inflation presented in chapter 6, the mode equation for $A_\lambda(\eta, k)$ includes the effects of a conductive medium, and reads

$$A_\lambda''(\eta, k) + \sigma_E \Theta A_\lambda'(\eta, k) + \left[k^2 - 2\lambda k a H \left(\xi + \frac{\sigma_B \Theta}{2H} \right) \right] A_\lambda(\eta, k) = 0, \quad (\text{A.11})$$

with electric and magnetic conductivities σ_E and σ_B , and $\Theta(\eta, k)$ a step function in η which we specify further below (see chapter 6 for the physical interpretation of $\Theta(\eta, k)$). Applying the substitution

$$A_\lambda(\eta, k) = \sqrt{\Delta_k(\eta)} f_\lambda(\eta, k), \quad \Delta_k(\eta) = \exp \left(- \int_{-\infty}^{\eta} a(\eta') \sigma_E(\eta') \Theta(\eta', k) d\eta' \right), \quad (\text{A.12})$$

one finds that the mode functions $f_\lambda(\eta, k)$ evolve according to (c.f., Eq. 6.40)

$$f_\lambda''(\eta, k) + \left[k^2 - 2\lambda k \xi_{\text{eff}} a H - \frac{s_E + s_E^2}{k^2} a^2 H^2 \right] f_\lambda(\eta, k) = 0. \quad (\text{A.13})$$

Here, we assumed that σ_E and σ_B are constant, and we introduced ξ_{eff} and s_E as defined in Eqs. (6.36) and (6.41). In de Sitter space, i.e., for $\eta = -1/(aH)$, Eq. (A.13) is also in the form of Eq. (A.1) with $\kappa = -i\lambda\xi_{\text{eff}}$ and $\mu = 1/2 + s_E$. Thus, $A_\lambda(\eta, k)$ is generally given by

$$A_\lambda(\eta, k) = \sqrt{\frac{\Delta_k(\eta)}{2k}} e^{\frac{\lambda\pi\xi_{\text{eff}}}{2}} \left[C_{\lambda,k} W_{-i\lambda\xi_{\text{eff}}, \frac{1}{2} + s_E}(2ik\eta) + D_{\lambda,k} M_{-i\lambda\xi_{\text{eff}}, \frac{1}{2} + s_E}(2ik\eta) \right], \quad (\text{A.14})$$

where $C_{\lambda,k}$ and $D_{\lambda,k}$ are time-independent coefficients.

In chapter 6, we have presented two choices for $\Theta(\eta, k)$. The simple version is that the function is constant for all η and k ; $\Theta(\eta, k) = 1$. In this case, the mode function $A_\lambda(\eta, k)$ obeys damped Bunch–Davies boundary conditions, Eq. (6.43). These boundary conditions fix $C_{\lambda,k} = 1$ and $D_{\lambda,k} = 0$ in Eq. (A.14), and one obtains

$$A_\lambda(\eta, k) = \sqrt{\frac{\Delta_k(\eta)}{2k}} e^{\frac{\lambda\pi\xi_{\text{eff}}}{2}} W_{-i\lambda\xi_{\text{eff}}, \frac{1}{2} + s_E}(2ik\eta). \quad (\text{A.15})$$

Alternatively, $\Theta(\eta, k)$ may be modelled as a step function as defined in Eq. (6.49). For the purposes of this appendix, this step function may be conveniently written in the form

$$\Theta(\eta, k) = \theta(\eta - \eta^*(k)), \quad (\text{A.16})$$

where $\eta^*(k)$ denotes a fixed moment in conformal time for every wavenumber k . This entails that a gauge mode function $A_\lambda(\eta, k)$ with a fixed wavenumber k solves Eq. (A.9) for $\eta < \eta^*(k)$, while solving Eq. A.11 with $\Theta(\eta, k) = 1$ for $\eta > \eta^*(k)$.

In the domain $\eta < \eta^*(k)$, $A_\lambda(\eta, k)$ may obey the conventional Bunch–Davies boundary condition, $A_\lambda(\eta, k) \rightarrow e^{-ik\eta}/\sqrt{2k}$ for $-k\eta \rightarrow \infty$ and the mode function takes the form of the Whittaker- W function in Eq. (A.10). In the domain $\eta > \eta^*(k)$, the expression for $A_\lambda(\eta, k)$ instead follows from Eq. (A.14). The coefficients $C_{\lambda,k}$ and $D_{\lambda,k}$ may then be determined by matching Eq. (A.14) and (A.10) at $\eta^*(k)$ to ensure that the physical mode function $A_\lambda(\eta, k)$ is smooth. This matching yields the following linear equations

$$W = C_{\lambda,k} \tilde{W} + D_{\lambda,k} \tilde{M}, \quad (\text{A.17a})$$

$$W' = C_{\lambda,k} \left[\tilde{W}' - i \frac{s_E}{2k\eta^*} \tilde{W} \right] + D_{\lambda,k} \left[\tilde{M}' - i \frac{s_E}{2k\eta^*} \tilde{M} \right], \quad (\text{A.17b})$$

where we have introduced the notation

$$W = W_{-i\lambda\xi, \frac{1}{2}}(2ik\eta^*), \quad \tilde{W} = W_{-i\lambda\xi_{\text{eff}}, \frac{1}{2} + s_E}(2ik\eta^*), \quad \tilde{M} = M_{-i\lambda\xi_{\text{eff}}, \frac{1}{2} + s_E}(2ik\eta^*), \quad (\text{A.18})$$

and where a prime denotes the derivative of a Whittaker function with respect to its argument, i.e., $W'_{\kappa,\mu}(z) = \frac{d}{dz} W_{\kappa,\mu}(z)$. Eq. (A.17) can be solved using Eqs. (A.6) to (A.8);

$$C_{\lambda,k} = -i \frac{\Gamma(1 + s_E + i\lambda\xi_{\text{eff}})}{(2k\eta^*)\Gamma(2 + 2s_E)} \left[W_{+1} \tilde{M} + (1 + s_E - i\lambda\xi_{\text{eff}}) W \tilde{M}_{+1} + (s_E + i\lambda s_B) W \tilde{M} \right], \quad (\text{A.19a})$$

$$D_{\lambda,k} = i \frac{\Gamma(1 + s_E + i\lambda\xi_{\text{eff}})}{(2k\eta^*)\Gamma(2 + 2s_E)} \left[W_{+1} \tilde{W} - W \tilde{W}_{+1} + (s_E + i\lambda s_B) W \tilde{W} \right]. \quad (\text{A.19b})$$

Here, the subscript +1 denotes the corresponding Whittaker function with the index κ increased by one unit, e.g., $W_{+1} = W_{1-i\lambda\xi, \frac{1}{2}}(2ik\eta^*)$. Also, we write $s_B = \xi_{\text{eff}} - \xi$.

Appendix B

Numerical determination of mode functions and power spectra

This appendix is partially adapted from the appendix in Ref. [4]. The detailed contributions of the author to this article is outlined at the beginning of this thesis.

As we have seen in chapters 4, the gradient-expansion formalism (GEF) allows us access to the dynamical evolution of the homogeneous and isotropic inflationary background. Based on these dynamics, we can then determine the spectral evolution of a variety of quantum fields during inflation, for instance that of the gauge field, the curvature perturbation, and the tensor metric perturbation. In this appendix, we provide details on the numerical strategies that we employ in the main text to compute the spectra for these fields based on pre-determined inflationary background dynamics from the GEF.

Specifically, in what follows, we assume that the evolution of the inflationary background is known starting from some initial time η_{ini} until some final time η_{end} with η denoting conformal time. This means, we assume that the typical background quantities like the Hubble rate, H , the scale factor a etc. are known functions of η . Alternatively to conformal time, we also use cosmic time t , or e-folds N which we define such that $N(\eta_{\text{ini}}) = 0$. Moreover, we implicitly assume a phase of accelerated expansion which is approximately described by de Sitter space such that $\eta \simeq -1/\mathcal{H}(\eta)$ with $\mathcal{H}$ being the comoving Hubble scale.

B.1 Solving a homogeneous mode equation

In multiple instances throughout this thesis, we have encountered the following generic situation: we want to determine the spectrum of a quantum field X_k that is given by mode functions $X(\eta, k)$ with $k = |\mathbf{k}|$, and these mode functions (i) evolve according to a homogeneous mode equation of the type

$$X''(\eta, k) + p_k(\eta)X'(\eta, k) + q_k(\eta)X(\eta, k) = 0, \quad (\text{B.1})$$

with real coefficients $p_k(\eta)$ and $q_k(\eta)$ depending on the inflationary background dynamics and k ; and (ii) evolve as positive-frequency plane waves at asymptotic early times $-k\eta \rightarrow \infty$. In these cases, the spectrum of X_k can be determined by numerically solving Eq. (B.1) for a range of wavenumbers k starting from some initial time η_{ini} until the end of inflation η_{end} . In this section, we outline a general strategy by which this can be achieved.

Specifically, this problem has appeared in two variations in the main text: (a) the gauge field in axion inflation (cf. Sec. 3.2) whose mode functions, $A_\lambda(\eta, k)$ evolve according to the homogeneous mode equation given by Eq. (3.19) (no Schwinger effect) or Eq. (6.35) (with the Schwinger effect); (b) the vacuum part of the curvature and tensor metric perturbation (cf. chapters 8 and 9) whose mode equations are respectively given by Eqs. (8.23) and (9.15).

Assumptions and Notation

Let us first mention restrictions on the form of the mode equation (B.1) which need to be met for our strategy to work. In particular, these requirements are necessary to ensure that Eq. (B.1) indeed admits an asymptotic solution $X(\eta, k) \propto e^{-ik\eta}$ for modes k that are deep inside the Hubble horizon, as these boundary conditions are later needed for us to define initial conditions for a mode $X(\eta, k)$ with fixed k . Note that all these requirements are fulfilled for the mode equations for which we use this strategy in the main text, and we only state them here for the sake of clarity and to establish some useful notation.

To discuss these restrictions, we need to rewrite Eq. (B.1) such that it takes the form of a harmonic oscillator with a time-dependent frequency. This can be achieved via the introduction of a rescaled mode function $Y(\eta, k)$ defined as

$$X(\eta, k) \propto \sqrt{\Delta_k(\eta)} Y(\eta, k), \quad \text{with} \quad \Delta_k(\eta) \equiv \exp\left(-\int_{\eta_{\text{ini}}}^{\eta} p_k(\eta') d\eta'\right), \quad (\text{B.2})$$

such that $Y(\eta, k)$ obeys

$$Y''(\eta, k) + \omega_k^2(\eta) Y(\eta, k) = 0, \quad \text{with} \quad \omega_k^2(\eta) \equiv \left(q_k - \frac{1}{2}p_k' - \frac{1}{4}p_k^2\right). \quad (\text{B.3})$$

The restrictions we consider are imposed on the frequency ω_k in this equation, which needs to behave as $\omega_k(\eta)^2 \simeq k^2$ for $-k/\mathcal{H}(\eta) \gg 1$ to ensure that we can indeed impose Bunch–Davies boundary conditions on $Y(\eta, k)$.

For this to be possible, a sufficient condition is that the coefficients $p_k(\eta)$ and $q_k(\eta)$ in Eq. (B.1) should behave as

$$q_k(\eta) \simeq k^2, \quad p_k(\eta) \sim \mathcal{H}(\eta), \quad \text{for} \quad -k/\mathcal{H}(\eta) \gg 1. \quad (\text{B.4})$$

and our strategy implicitly assumes that this condition is satisfied.

Moreover, our strategy requires some additional structure for the frequency term ω_k for reasons that become clear later on. Specifically, our numerical strategy relies on the existence of a characteristic momentum scale k_c in the mode equation Eq. (B.1) which ensures

$$k_c(\eta) \gtrsim \mathcal{H}(\eta), \quad \text{such that} \quad \omega_k(\eta)^2 \simeq k^2 \quad \text{for} \quad k \gg k_c(\eta). \quad (\text{B.5})$$

If this characteristic scale exists, it means that we can always write the boundary condition for $Y(\eta, k)$ as¹

$$Y(\eta, k) \simeq \frac{1}{\sqrt{2k}} e^{-ik\eta} \quad \text{for} \quad k \gg k_c(\eta) \gtrsim \mathcal{H}(\eta). \quad (\text{B.6})$$

Looking back at the particular mode equations appearing in the main text, the scale k_c is either the instability scale k_h for the gauge field mode functions in axion inflation, or simply the Hubble scale $\mathcal{H}$ for the mode functions of the curvature and tensor metric perturbations. For other cases, this scale needs to be appropriately identified.

¹This normalization of $Y(\eta, k)$ requires an appropriate choice of the proportionality factor in Eq. (B.2). This proportionality factor is unimportant for our present discussion, but we emphasize that it cannot be arbitrary, and needs to be chosen appropriately depending on the concrete physical system. For instance, in the case of the mode functions for vacuum curvature perturbations, we need that the vacuum mode functions are normalized as in, e.g., Eq. (8.83).

The strategy

Having discussed these requirements, the strategy for solving Eq. (B.1) simply involves writing Eq. (B.1) in a form that is more suitable for numerical integration, and defining a simple algorithm for initializing modes in their asymptotic boundary state.

First, we rewrite Eq. (B.1) as a set of coupled linear differential equations. To this end, the following choice of variables have proven to be useful

$$f_k(\eta) \equiv \sqrt{2k}Y(\eta, k), \quad g_k(\eta) \equiv \sqrt{\frac{2}{k}} \left(Y'(\eta, k) - \frac{1}{2}p'_k(\eta)Y(\eta, k) \right). \quad (\text{B.7})$$

These functions simply evolve according to

$$f'_k = kg_k + \frac{1}{2}p_k f_k, \quad g'_k = -\frac{q_k}{k}f_k + \frac{1}{2}p_k g_k, \quad (\text{B.8})$$

and they are asymptotically given by

$$f_k(\eta) \simeq e^{-ik\eta}, \quad g_k(\eta) \simeq -ie^{-ik\eta}, \quad \text{for } k \gg k_c(\eta), \quad (\text{B.9})$$

where we used Eq. (B.6) and Eq. (B.4).

For a fixed wavenumber k , we then compute the evolution of f_k and g_k numerically by solving Eq. (B.8) in terms of cosmic time t . To initialize f_k and g_k , we consider that their evolution is well-approximated by Eq. (B.9) for times $t \leq t_k$, with t_k implicitly defined through $k = 10^{5/2}k_c(t_k)$. Therefore, we can explicitly use $f_k(t_k) = e^{-ik\eta(t_k)}$ and $g_k(t_k) = -ie^{-ik\eta(t_k)}$ as initial conditions and evolve these functions starting from t_k onward. This initialization procedure has proven to be useful because f_k and g_k are analytical simple, but rapidly oscillating functions as long as $k \gg k_c$. Numerically evolving such functions forward in time can be extremely costly, and by only computing f_k and g_k for times $t \geq t_k$ we can significantly reduce this cost. In particular, this initialization procedure requires the existence of a characteristic scale k_c as given in Eq. (B.5).

Finally, this approach can be readily applied to determine $X(k, \eta)$ for any wavenumber

$$k \in [10^{5/2}k_c(t_{\text{ini}}), 10^{5/2}k_c(t_{\text{end}})] \quad \text{with} \quad t_{\text{ini/end}} = t(\eta_{\text{ini/end}}), \quad (\text{B.10})$$

by using Eqs. (B.2) and (B.7) to relate f_k to $X(k, \eta)$ (cf. footnote 1 regarding the constant of proportionality in Eq. (B.2)).

B.2 Retarded Green function

In chapters 8 and 9, we have noted the necessity for determining the retarded Green function for a mode equation as in Eq. (B.1), i.e., the function defined by

$$[\partial_\eta^2 + p_k(\eta)\partial_\eta + q_k(\eta)] G_k(\eta, \tau) = \delta(\eta - \tau), \quad (\text{B.11a})$$

$$G_k(\eta, \tau \geq \eta) \equiv 0, \quad \left. \frac{\partial}{\partial \eta} G_k(\eta, \tau) \right|_{\tau \rightarrow \eta} = 1. \quad (\text{B.11b})$$

Here, we present a simple numerical strategy that we can use to determine $G_k(\eta, \tau)$. Note that, we assume that p_k and q_k behave as described in the Sec. B.1, i.e., they asymptotically behave as in Eq. (B.4), and we suppose the existence of a critical scale $k_c(\eta)$ as in Eq. (B.5).

First, we can note that the retarded Green function $G_k(\eta, \tau)$ can be directly written in terms of the mode function $X(\eta, k)$ that solves Eq. (B.1):

$$G_k(\eta, \tau) = \frac{\text{Im}[X^*(\eta, k)X(\tau, k)]}{\text{Im}[X(\tau, k)X^*(\tau, k)]}\theta(\eta - \tau). \quad (\text{B.12})$$

This is possible since $X(\eta, k)$ and its complex conjugate constitute two linearly independent solutions to Eq. (B.1). Therefore, one could, in principle, use the approach in the previous section to determine numerical solutions for $X(\eta, k)$ and use these to directly evaluate Eq. (B.12). However, this strategy proves impractical, because the term in the denominator in Eq. (B.12) suffers from numerical cancellation as $\eta \rightarrow \tau$. An alternative, more stable approach to obtain G_k is to write Eq. (B.12) as

$$G_k(\eta, \tau) = \frac{B_k^\eta(\tau)}{k}\theta(\eta - \tau). \quad (\text{B.13})$$

and to directly evolving the function $B_k^\eta(\tau)$ in time for $\tau \leq \eta$ and fixed values of k and η .

To do so, we define a second function $C_k^\eta(\tau)$ alongside $B_k^\eta(\tau)$, writing the two as

$$B_k^\eta(\tau) \equiv k \frac{\text{Im}[X^*(\eta, k)X(\tau, k)]}{\text{Im}[X(\tau, k)X^*(\tau, k)]}, \quad C_k^\eta(\tau) \equiv \frac{\text{Im}[X^{*'}(\eta, k)X(\tau, k)]}{\text{Im}[X(\tau, k)X^*(\tau, k)]}. \quad (\text{B.14})$$

From Eq. (B.1), it can be readily derived that these two functions evolve according to

$$\partial_\tau B_k^\eta = kC_k^\eta + p_k B_k^\eta, \quad \partial_\tau C_k^\eta = -\frac{q_k}{k} B_k^\eta. \quad (\text{B.15})$$

where we emphasize that the derivative ∂_τ acts on the argument τ of $B_k^\eta(\tau)$ and $C_k^\eta(\tau)$, not the superscript η .

In practice, we can then obtain B_k^η directly by solving Eq. (B.15) backwards in time starting from $\tau \rightarrow \eta$, where we have particularly simple initial conditions given by

$$B_k^\eta(\eta) = 0, \quad \text{and} \quad C_k^\eta(\eta) = 1. \quad (\text{B.16})$$

These directly follow from the definitions of these functions in Eq. (B.15) and the fact that the imaginary part of a real scalar vanishes. In particular, we determine the evolution of B_k^η using cosmic time t as an integration variable for Eq. (B.15). For the sake of efficiency, we only evolve B_k^η until a time t_k defined by $k = 10^{5/2}k_c(t_k)$ as in the previous section. For times $t \leq t_k$ we instead use that $B_k^\eta(\tau)$ can be written as

$$B_k^\eta(\tau) \simeq \sqrt{\frac{\Delta_k(\eta)}{\Delta_k(\tau)}} \text{Im}[f_k^*(\eta)e^{-ik\tau}], \quad \text{for } k \gg k_c(\tau), \quad (\text{B.17})$$

which directly follows from Eqs. (B.2), (B.7) and (B.14). Thus, only the value $f_k^*(\eta)$ is needed to approximately describe the behavior of $B_k^\eta(\tau)$ for $t \leq t_k$, and these can be readily determined using the strategy described in Sec. B.1.

As in the last section, Eq. (B.13) implies that we can then determine $G_k(\eta, \tau)$ in Eq. (B.13) for any wavenumber $k \in [10^{5/2}k_c(t_{\text{ini}}), 10^{5/2}k_c(t_{\text{end}})]$, and for any η and τ with $\eta_{\text{ini}} \leq \tau \leq \eta \leq \eta_{\text{end}}$.

B.3 Induced scalar and tensor power spectra

In chapters 8 and 9, we present numerical results for the induced scalar and tensor power spectra for axion inflation that we determine by integrating their respective integral expressions in Eqs. (8.39) and (9.25). Here, we briefly summarize how we compute these integrals in practice. The approach we present here is inspired by the discussion in the appendices of Ref. [83].

Both integrals in question can be schematically written in the form

$$\mathcal{P}_X(\eta, k) \propto \int d^3\mathbf{p} \sum_{\alpha, \beta = \pm 1} \left| \int_{-\infty}^0 d\tau G_k(\eta, \tau) \mathcal{K}(\tau; p, q; \alpha, \beta) \right|^2 \quad (\text{B.18})$$

where we have three momenta $p = |\mathbf{p}|$, $k = |\mathbf{k}|$ and $q = |\mathbf{k} - \mathbf{p}|$, a retarded Green function G_k , and the kernel $\mathcal{K}$ depends on gauge-field mode functions $A_\alpha(\tau, p)$ and $A_\beta(\tau, q)$. Moreover, although not explicitly apparent in Eq. (B.18), $\mathcal{P}_X(\eta, k)$ is UV divergent, and we regularize it by enforcing that the momenta inside the integral satisfy $q, p \leq k_h(\eta)$ with k_h the instability scale for the gauge field during axion inflation (see Eqs. (3.24) and (6.56)).

To compute $\mathcal{P}_X(\eta, k)$ for a definite time $\eta \leq \eta_{\text{end}}$ and a fixed momentum k , we first use the approach in Sec. B.2 to numerically determine the Green function $G_k(\eta, \tau)$ for $\eta_{\text{ini}} \leq \tau \leq \eta$. Similarly, we compute the spectrum of mode functions $A_\lambda(\tau, \tilde{k})$ for a discrete set of wavenumbers $\tilde{k} \in [k_{\text{IR}}, k_h(\eta)]$ using the strategy in Sec. B.1. Here, $k_{\text{IR}} = 10^{5/2} k_h(\eta_{\text{ini}})$ is the smallest wavenumber that we can access given that we know the inflationary dynamics from η_{ini} onward, and $k_h(\eta)$ is the largest momentum that enters in Eq. (B.18).

Next, we introduce the integration variables suggested in Ref. [83],

$$A = \frac{1}{2k}(p + q), \quad B = \frac{1}{2k}(p - q), \quad (\text{B.19})$$

and re-express Eq. (B.18) as an integral over e -folds N . This yields

$$\begin{aligned} \mathcal{P}_X^{\text{ind}}(k, \eta) &\propto (4\pi k^2) \int_{A_{\text{min}}}^{A_{\text{max}}} dA \int_{-B_0(A)}^{B_0(A)} dB |A - B| \\ &\times \sum_{\alpha, \beta = \pm 1} \left| \int_0^{N(\eta)} dN' B_k^\eta(\tau(N')) \mathcal{K}(\tau(N'); k(A + B), k(A - B); \alpha, \beta) \right|^2. \end{aligned} \quad (\text{B.20})$$

To arrive at this expression, we have used Eq. (B.13) to express G_k through B_k^η and the integration boundaries in Eq. (B.21a) are defined through

$$A_{\text{min}} = \max\left(A_{\text{IR}}, \frac{1}{2}\right), \quad A_{\text{max}} = A_{\text{UV}}, \quad (\text{B.21a})$$

$$B_0(A) = \begin{cases} A - A_{\text{IR}}, & \text{for } A_{\text{min}} \leq A \leq \frac{1}{2} + A_{\text{IR}}; \\ \frac{1}{2}, & \text{for } \frac{1}{2} + A_{\text{IR}} < A \leq A_{\text{UV}} - \frac{1}{2}; \\ A_{\text{UV}} - A, & \text{for } A_{\text{UV}} - \frac{1}{2} < A \leq A_{\text{UV}}, \end{cases} \quad (\text{B.21b})$$

with $A_{\text{IR}} = k_{\text{IR}}/k$ and $A_{\text{UV}} = k_h(\eta)/k$. These integration boundaries are needed to ensure we only integrate over momenta $p, q \in [k_{\text{IR}}, k_h(\eta)]$.

Finally, the integral in Eq. (B.21a) can be computed by first determining the inner integral over N using the trapezoid rule, and then performing the integral over A and B by first interpolating the spectrum of mode functions $A_\lambda(\tau, \tilde{k})$ for fixed $\tau(N)$ and subsequent quadrature integration.

Appendix C

General expressions for electric and magnetic conductivities

In this appendix, we give general expressions for the electric and magnetic conductivities σ_E and σ_B that appear in our effective description of Schwinger pair production during axion inflation presented in chapter 6. Therein, we have presented three possible representations for these conductivities: the *electric picture*, with $\sigma_E \neq 0$ and $\sigma_B = 0$, the *magnetic picture*, with $\sigma_E = 0$ and $\sigma_B \neq 0$, and the *mixed picture*, with both $\sigma_E \neq 0$ and $\sigma_B \neq 0$.

The fundamental result which motivates the three representations of σ_E and σ_B is the expressions for the Schwinger-induced current $\mathbf{J}$ for a configuration of constant and collinear electric and magnetic fields in de Sitter space, as derived in Refs. [118, 397, 398, 402]. For a single species of charge carriers with mass m and charge Q , the magnitude of this current is given by

$$|\mathbf{J}| = \frac{(e|Q|)^3}{6\pi^2 H} |\mathbf{E}||\mathbf{B}| \psi\left(\frac{\pi|\mathbf{B}|}{|\mathbf{E}|}\right) \exp\left(-\frac{\pi m^2}{e|Q\mathbf{E}|}\right), \quad (\text{C.1})$$

where e denotes the gauge coupling associated with the gauge charge Q , H is the Hubble rate, $\mathbf{E}$ and $\mathbf{B}$, respectively, denote electric and magnetic fields, and $\psi(x) = \coth x$ for a fermion or $\psi(x) = 1/(2 \sinh x)$ for a charged scalar.

Expressions for the electric and magnetic picture

Both the electric and the magnetic picture directly follow from Eq. (C.1). Assuming multiple, non-interacting charge carrier species χ_i with respective masses m_i and charges Q_i , each species will constitute a charged current $\mathbf{J}_i$ given by Eq. (C.1). The total induced current from all these fields is given by the sum over the individual charged currents; $\mathbf{J} = \sum_i \mathbf{J}_i$. For collinear electric and magnetic fields, Eq. (C.1) suggests that the same holds for their magnitudes; $|\mathbf{J}| = \sum_i |\mathbf{J}_i|$.

We now make the Ansatz $\mathbf{J} = \sigma_E \mathbf{E}$. Following the above considerations, this yields the general expression for σ_E in the electric picture:

$$\sigma_E = \frac{e^3}{6\pi^2 H} \sqrt{\langle \mathbf{B}^2 \rangle} \sum_i |Q_i|^3 \psi_i \left(\pi \sqrt{\frac{\langle \mathbf{B}^2 \rangle}{\langle \mathbf{E}^2 \rangle}} \right) \exp \left(-\frac{\pi m_i^2}{e|Q_i| \sqrt{\langle \mathbf{E}^2 \rangle}} \right). \quad (\text{C.2})$$

Here, we have replaced the magnitudes of $\mathbf{E}$ and $\mathbf{B}$ by their respective RMS values, $\langle \mathbf{E}^2 \rangle$ and $\langle \mathbf{B}^2 \rangle$ according to the assumptions behind our effective model in 6. Additionally, for each charge-carrier species χ_i , we, respectively, have $\psi_i(x) = \coth x$ or $\psi_i(x) = 1/(2 \sinh x)$ for fermions and charged scalars.

Similarly, we can derive the general expression for the magnetic picture by making the Ansatz $\mathbf{J} = \sigma_E \mathbf{B}$. This yields

$$\sigma_B = \frac{e^3}{6\pi^2 H} \text{sign}(\langle \mathbf{E} \cdot \mathbf{B} \rangle) \sqrt{\langle \mathbf{E}^2 \rangle} \sum_i |Q_i|^3 \psi_i \left(\pi \sqrt{\frac{\langle \mathbf{B}^2 \rangle}{\langle \mathbf{E}^2 \rangle}} \right) \exp \left(-\frac{\pi m_i^2}{e|Q_i| \sqrt{\langle \mathbf{E}^2 \rangle}} \right), \quad (\text{C.3})$$

where $\text{sign} \langle \mathbf{E} \cdot \mathbf{B} \rangle$ accounts for the fact that Eq. (C.1) is also valid for $\mathbf{E}$ and $\mathbf{B}$ in an anti-collinear configuration.

Expressions for the mixed picture

The mixed-picture representation of σ_E and σ_B is valid beyond the assumption of collinear electric and magnetic fields. However, in a frame where $\mathbf{E}$ and $\mathbf{B}$ are collinear, Eq. (C.1) is still valid. Moreover, for multiple charge carriers χ_i , the magnitudes of their individual charge currents $\mathbf{J}_i$ are still additive in the collinear frame, i.e., $|\mathbf{J}'| = \sum_i |\mathbf{J}'_i|$, with prime (') denoting quantities in the collinear frame. Following the procedure in chapter 6 for boosting to the comoving frame, where $\mathbf{E}$ and $\mathbf{B}$ are not collinear, and using Eq. (6.29), the general expressions for σ_E and σ_B are then given by

$$\sigma_E = \frac{e^3}{6\pi^2 H} \frac{|\langle \mathbf{E} \cdot \mathbf{B} \rangle|}{\sqrt{\Sigma}} \sqrt{\frac{\langle \mathbf{E}^2 - \mathbf{B}^2 \rangle + \Sigma}{\langle \mathbf{E}^2 + \mathbf{B}^2 \rangle + \Sigma}} \sum_i \mathfrak{J}_i, \quad (\text{C.4a})$$

$$\sigma_B = \frac{e^3}{6\pi^2 H} \frac{\langle \mathbf{E} \cdot \mathbf{B} \rangle}{\sqrt{\Sigma}} \sqrt{\frac{\langle \mathbf{B}^2 - \mathbf{E}^2 \rangle + \Sigma}{\langle \mathbf{E}^2 + \mathbf{B}^2 \rangle + \Sigma}} \sum_i \mathfrak{J}_i, \quad (\text{C.4b})$$

where

$$\Sigma = \sqrt{\langle \mathbf{E}^2 - \mathbf{B}^2 \rangle^2 + 4\langle \mathbf{E} \cdot \mathbf{B} \rangle^2}, \quad (\text{C.5})$$

and

$$\mathfrak{J}_i = |Q_i|^3 \psi_i \left(\pi \sqrt{\frac{\langle \mathbf{B}^2 - \mathbf{E}^2 \rangle + \Sigma}{\langle \mathbf{E}^2 - \mathbf{B}^2 \rangle + \Sigma}} \right) \exp \left(-\frac{\pi m_i^2 \sqrt{2}}{e|Q_i|(\langle \mathbf{B}^2 - \mathbf{E}^2 \rangle + \Sigma)^{1/2}} \right). \quad (\text{C.6})$$

Appendix D

Equation of motion for ζ and scalar source functions

This appendix is adapted from the appendices in Ref. [2]. The detailed contributions of the author to this article is outlined at the beginning of this thesis.

In this appendix, we give supplementary information to chapter 8. In the first section, we present a detailed derivation of the master equation for the curvature perturbation variable ζ in Eq. (8.17). The second section gives the full expressions for the source terms L and K entering the scalar power spectrum and bispectrum in Eqs. (8.39b) and (8.42). Regarding notation, this appendix is consistent with chapter 8.

D.1 Deriving the equation of motion for ζ

In what follows, we discuss how one may arrive at the equation of motion for the curvature perturbation variable ζ in Eq. (8.17).¹

To start, we derive three auxiliary equations from the linearized Einstein equations (8.13).

- (i) By substituting $\Psi = \Phi - 2F_\sigma$ into Eq. (8.13a) and solving for $\delta\phi'$ we obtain

$$\delta\phi' = \frac{2M_{\text{P}}^2}{\varphi'} \left\{ \left[-k^2 + \frac{\varphi'^2}{2M_{\text{P}}^2} - \frac{a^2}{M_{\text{P}}^2} \langle \mathbf{B}^2 \rangle \right] \Phi - \frac{1}{2M_{\text{P}}^2} \left[3\mathcal{H}\varphi' + a^2 V_{,\phi}(\varphi) \right] \delta\phi - F_\rho - 3\mathcal{H}F_q - \left[\frac{\varphi'^2}{M_{\text{P}}^2} + \frac{a^2}{M_{\text{P}}^2} \langle \mathbf{E}^2 \rangle \right] F_\sigma \right\}. \quad (\text{D.1})$$

- (ii) Similarly, we substitute $\Psi = \Phi - 2F_\sigma$ into Eq. (8.13b), solving for Φ' to find

$$\Phi' = -\mathcal{H}\Phi + \frac{\varphi'}{2M_{\text{P}}^2} \delta\phi + F_q + 2\mathcal{H}F_\sigma. \quad (\text{D.2})$$

- (iii) Next, we seek to determine the combination $(\Phi' + \mathcal{H}\Psi)'$. First, we use Eq. (8.13c) and $(\Phi' + \mathcal{H}\Psi)' = \Phi'' + \mathcal{H}\Psi' + \mathcal{H}'\Psi$, to obtain

$$(\Phi' + \mathcal{H}\Psi)' = \frac{a^2}{2M_{\text{P}}^2} \left\{ \frac{\varphi'}{a^2} \delta\phi' - V_{,\phi}(\varphi) \delta\phi - \left[\frac{\varphi'^2}{a^2} + \frac{1}{3} \langle \mathbf{E}^2 \rangle \right] \Psi + \frac{1}{3} \left[\langle \mathbf{E}^2 \rangle + 2\langle \mathbf{B}^2 \rangle \right] \Phi \right\} - 2\mathcal{H}\Phi' - (\mathcal{H}' + \mathcal{H}^2)\Psi + \frac{1}{3}F_p - \frac{2}{3}k^2 F_\sigma. \quad (\text{D.3})$$

¹We implicitly work in Fourier space, so the Laplace operator appearing in all equations in the main text is replaced as $\Delta \rightarrow -k^2$ with $k = |\mathbf{k}|$, the absolute value of a Fourier-mode's wavevector.

Substituting $\Psi = \Phi - 2F_\sigma$ and Eq. (D.2), one gets

$$(\Phi' + \mathcal{H}\Psi)' = \left[-k^2 + \frac{\varphi'^2}{2M_{\text{P}}^2} + \frac{a^2}{3M_{\text{P}}^2} (\langle \mathbf{E}^2 \rangle - \langle \mathbf{B}^2 \rangle) \right] \Phi - \frac{1}{2M_{\text{P}}^2} \left[5\mathcal{H}\varphi' + 2a^2 V_{,\phi}(\varphi) \right] \delta\phi \\ - \frac{2}{3} F_\rho - 5\mathcal{H}F_q - \left[\frac{2}{3} k^2 + \frac{\varphi'^2}{M_{\text{P}}^2} + \frac{2a^2}{3M_{\text{P}}^2} (2\langle \mathbf{E}^2 \rangle + \langle \mathbf{B}^2 \rangle) \right] F_\sigma. \quad (\text{D.4})$$

With these equations in place, we now turn toward the curvature perturbation variable ζ .

(iv) The expression for ζ in Eq. (8.16) can be written as

$$\zeta = M_{11}\Phi + M_{12}\delta\phi + c_1^{(\rho)}F_\rho + c_1^{(q)}F_q + c_1^{(\sigma)}F_\sigma, \quad (\text{D.5})$$

with

$$M_{11} = 1, \quad M_{12} = \frac{\mathcal{H}}{\mathcal{H}^2 - \mathcal{H}'} \frac{\varphi'}{2M_{\text{P}}^2}, \quad (\text{D.6})$$

and

$$c_1^{(\rho)} = c_1^{(\sigma)} = 0, \quad c_1^{(q)} = \frac{\mathcal{H}}{\mathcal{H}^2 - \mathcal{H}'}. \quad (\text{D.7})$$

(v) We then determine a similar form for ζ' by taking the derivative of

$$\zeta = \Phi + \frac{\mathcal{H}}{\mathcal{H}^2 - \mathcal{H}'} (\Phi' + \mathcal{H}\Psi) \quad (\text{D.8})$$

and using Eqs. (D.2) and (D.4). This yields

$$\zeta' = M_{21}\Phi + M_{22}\delta\phi + c_2^{(\rho)}F_\rho + c_2^{(q)}F_q + c_2^{(\sigma)}F_\sigma, \quad (\text{D.9})$$

with

$$M_{21} = -\frac{\mathcal{H}}{\mathcal{H}^2 - \mathcal{H}'} \left[k^2 + \frac{2a^2}{3M_{\text{P}}^2} \langle \mathbf{B}^2 \rangle \right], \quad (\text{D.10a})$$

$$M_{22} = -\frac{a^2}{3M_{\text{P}}^4} \frac{\mathcal{H}}{(\mathcal{H}^2 - \mathcal{H}')^2} \left\{ \frac{\beta\varphi'^2}{2M_{\text{P}}} \langle (\mathbf{E} \cdot \mathbf{B}) \rangle + (\langle \mathbf{E}^2 \rangle + \langle \mathbf{B}^2 \rangle) \left[\mathcal{H}\varphi' + a^2 V_{,\phi}(\varphi) \right] \right\}, \quad (\text{D.10b})$$

and

$$c_2^{(\rho)} = -\frac{2}{3} \frac{\mathcal{H}}{\mathcal{H}^2 - \mathcal{H}'}, \quad c_2^{(\sigma)} = -\frac{2}{3} \frac{\mathcal{H}}{\mathcal{H}^2 - \mathcal{H}'} \left[k^2 + \frac{a^2}{M_{\text{P}}^2} \langle \mathbf{E}^2 \rangle \right], \quad (\text{D.11a})$$

$$c_2^{(q)} = 1 + \left(\frac{\mathcal{H}}{\mathcal{H}^2 - \mathcal{H}'} \right)' - 5 \frac{\mathcal{H}^2}{\mathcal{H}^2 - \mathcal{H}'} \\ = \frac{\mathcal{H}a^2}{M_{\text{P}}^2(\mathcal{H}^2 - \mathcal{H}')^2} \left\{ \varphi' \left[V_{,\phi}(\varphi) - \frac{\beta}{3M_{\text{P}}} \langle (\mathbf{E} \cdot \mathbf{B}) \rangle \right] - \frac{2}{3} \mathcal{H} (\langle \mathbf{E}^2 \rangle + \langle \mathbf{B}^2 \rangle) \right\}. \quad (\text{D.11b})$$

(vi) Eqs. (D.5) and (D.9) form a set of linear equations relating the vector (ζ, ζ') to $(\Phi, \delta\phi)$ via a matrix M and the source functions F_α via vectors $c^{(\alpha)} = (c_1^{(\alpha)}, c_2^{(\alpha)})$. Inverting this equation

yields

$$\Phi = (M^{-1})_{11}\zeta + (M^{-1})_{12}\zeta' - \sum_{\alpha=\rho,\sigma,q} \left[(M^{-1})_{11}c_1^{(\alpha)} + (M^{-1})_{12}c_2^{(\alpha)} \right] F_\alpha, \quad (\text{D.12})$$

$$\delta\varphi = (M^{-1})_{21}\zeta + (M^{-1})_{22}\zeta' - \sum_{\alpha=\rho,\sigma,q} \left[(M^{-1})_{21}c_1^{(\alpha)} + (M^{-1})_{22}c_2^{(\alpha)} \right] F_\alpha, \quad (\text{D.13})$$

where M^{-1} denotes the inverse matrix of the matrix M .

(vii) We then take the time derivative of Eq. (D.9) to obtain

$$\zeta'' = M_{21}\Phi' + M'_{21}\Phi + M_{22}\delta\varphi' + M'_{22}\delta\varphi + \sum_{\alpha=\rho,\sigma,v} \left[c_2^{(\alpha)}F'_\alpha + (c_2^{(\alpha)})'F_\alpha \right]. \quad (\text{D.14})$$

Inserting Eqs. (D.2), (D.1), and the equations of motion for the sources Eq. (8.14), we can express ζ'' in terms of Φ , φ , and the sources F_ρ , F_q , F_σ , F'_σ , and $\delta_{\mathbf{E}\cdot\mathbf{B}}$:

$$\zeta'' = d^{(\Phi)}\Phi + d^{(\delta\varphi)}\delta\varphi + d^{(q)}F_q + d^{(\rho)}F_\rho + d_1^{(\sigma)}F_\sigma + d_2^{(\sigma)}F'_\sigma + d^{(\mathbf{E}\cdot\mathbf{B})}\delta_{\mathbf{E}\cdot\mathbf{B}}. \quad (\text{D.15})$$

The coefficients in Eq. (D.15) are given by

$$\begin{aligned} d^{(\Phi)} = M'_{21} - \mathcal{H}M_{21} - \frac{2M_{\text{P}}^2}{\varphi'} \left[k^2 + \frac{a^2}{M_{\text{P}}^2} \langle \mathbf{B}^2 \rangle - \frac{\varphi'^2}{2M_{\text{P}}^2} \right] M_{22} + \frac{a^2}{3M_{\text{P}}^2} (\langle \mathbf{E}^2 \rangle + 2\langle \mathbf{B}^2 \rangle) c_2^{(q)} \\ + \frac{a^2}{M_{\text{P}}^2} \left\{ \frac{\beta}{M_{\text{P}}} \frac{M_{\text{P}}^2}{\varphi'} \langle (\mathbf{E} \cdot \mathbf{B}) \rangle \left[k^2 + \frac{a^2}{M_{\text{P}}^2} \langle \mathbf{B}^2 \rangle - \frac{\varphi'^2}{2M_{\text{P}}^2} \right] \right. \\ \left. - \mathcal{H} \langle \mathbf{E}^2 \rangle + 2\langle (\mathbf{E} \cdot \text{rot } \mathbf{B}) \rangle \right\} c_2^{(\rho)}, \end{aligned} \quad (\text{D.16a})$$

$$\begin{aligned} d^{(\delta\varphi)} = M'_{22} + \frac{\varphi'}{2M_{\text{P}}^2} M_{21} - \left[3\mathcal{H} + \frac{a^2}{\varphi'} V_{,\phi}(\varphi) \right] M_{22} - \frac{\beta}{M_{\text{P}}} \frac{a^2}{2M_{\text{P}}^2} \langle (\mathbf{E} \cdot \mathbf{B}) \rangle c_2^{(q)} \\ + \frac{a^2}{2M_{\text{P}}^2} \left\{ \frac{\beta}{M_{\text{P}}} \langle (\mathbf{E} \cdot \mathbf{B}) \rangle \left[3\mathcal{H} + \frac{a^2}{\varphi'} V_{,\phi}(\varphi) \right] + \frac{\varphi'}{M_{\text{P}}^2} \langle \mathbf{E}^2 \rangle \right\} c_2^{(\rho)}, \end{aligned} \quad (\text{D.16b})$$

$$\begin{aligned} d^{(q)} = c_2^{(q)'} - 2\mathcal{H}c_2^{(q)} + M_{21} - \frac{6M_{\text{P}}^2\mathcal{H}}{\varphi'} M_{22} \\ - \left\{ k^2 - \frac{a^2}{M_{\text{P}}^2} \langle \mathbf{E}^2 \rangle - \frac{\beta}{M_{\text{P}}} \frac{3\mathcal{H}a^2}{\varphi'} \langle (\mathbf{E} \cdot \mathbf{B}) \rangle \right\} c_2^{(\rho)}, \end{aligned} \quad (\text{D.16c})$$

$$d^{(\rho)} = c_2^{(\rho)'} - \left\{ 2\mathcal{H} - \frac{\beta}{M_{\text{P}}} \frac{a^2}{\varphi'} \langle (\mathbf{E} \cdot \mathbf{B}) \rangle \right\} c_2^{(\rho)} + \frac{1}{3} c_2^{(q)} - \frac{2M_{\text{P}}^2}{\varphi'} M_{22}, \quad (\text{D.16d})$$

$$\begin{aligned} d_1^{(\sigma)} = c_2^{(\sigma)'} + 2\mathcal{H}M_{21} - \left[2\varphi' + \frac{2a^2}{\varphi'} \langle \mathbf{E}^2 \rangle \right] M_{22} - \left[\frac{2}{3} k^2 + \frac{a^2}{3M_{\text{P}}^2} (\langle \mathbf{E}^2 \rangle + 2\langle \mathbf{B}^2 \rangle) \right] c_2^{(q)} \\ + \frac{a^2}{M_{\text{P}}^2} \left\{ \frac{\beta}{M_{\text{P}}} \langle (\mathbf{E} \cdot \mathbf{B}) \rangle \left[2\varphi' + \frac{a^2}{\varphi'} \langle \mathbf{E}^2 \rangle \right] + 2[\mathcal{H} \langle \mathbf{E}^2 \rangle - \langle (\mathbf{E} \cdot \text{rot } \mathbf{B}) \rangle] \right\} c_2^{(\rho)}, \end{aligned} \quad (\text{D.16e})$$

$$d_2^{(\sigma)} = c_2^{(\sigma)} - \frac{a^2}{M_{\text{P}}^2} \langle \mathbf{E}^2 \rangle c_2^{(\rho)} = -\frac{2k^2}{3} \frac{\mathcal{H}}{\mathcal{H}^2 - \mathcal{H}'}, \quad (\text{D.16f})$$

$$d^{(\mathbf{E}\cdot\mathbf{B})} = -\frac{\beta}{M_{\text{P}}} \frac{a^2\varphi'}{2M_{\text{P}}^2} c_2^{(\rho)} = \frac{\beta}{M_{\text{P}}} \frac{a^2\varphi'}{3M_{\text{P}}^2} \frac{\mathcal{H}}{\mathcal{H}^2 - \mathcal{H}'}. \quad (\text{D.16g})$$

(viii) In a final step, one then uses Eqs. (D.12) and (D.13) to arrive at Eq. (8.17) from the main text

$$\zeta_{\mathbf{k}}'' + p_{\mathbf{k}} \zeta_{\mathbf{k}}' + q_{\mathbf{k}} \zeta_{\mathbf{k}} = r^{(q)} F_v + r^{(\rho)} F_\rho + r_1^{(\sigma)} F_\sigma + r_2^{(\sigma)} F'_\sigma + r^{(\mathbf{E} \cdot \mathbf{B})} \delta_{\mathbf{E} \cdot \mathbf{B}}, \quad (\text{D.17})$$

Here, we have again included the lower index $\mathbf{k}$ indicating the momentum of the Fourier mode. The coefficients in this equation are given by

$$p_{\mathbf{k}} = -(M^{-1})_{12} d^{(\Phi)} - (M^{-1})_{22} d^{(\delta\varphi)}, \quad (\text{D.18a})$$

$$q_{\mathbf{k}} = -(M^{-1})_{11} d^{(\Phi)} - (M^{-1})_{21} d^{(\delta\varphi)}, \quad (\text{D.18b})$$

$$r^{(q)} = d^{(q)} - \left[(M^{-1})_{11} c_1^{(q)} + (M^{-1})_{12} c_2^{(q)} \right] d^{(\Phi)} - \left[(M^{-1})_{21} c_1^{(q)} + (M^{-1})_{22} c_2^{(q)} \right] d^{(\delta\varphi)}, \quad (\text{D.18c})$$

$$r^{(\rho)} = d^{(\rho)} - \left[(M^{-1})_{11} c_1^{(\rho)} + (M^{-1})_{12} c_2^{(\rho)} \right] d^{(\Phi)} - \left[(M^{-1})_{21} c_1^{(\rho)} + (M^{-1})_{22} c_2^{(\rho)} \right] d^{(\delta\varphi)}, \quad (\text{D.18d})$$

$$r_1^{(\sigma)} = d_1^{(\sigma)} - \left[(M^{-1})_{11} c_1^{(\sigma)} + (M^{-1})_{12} c_2^{(\sigma)} \right] d^{(\Phi)} - \left[(M^{-1})_{21} c_1^{(\sigma)} + (M^{-1})_{22} c_2^{(\sigma)} \right] d^{(\delta\varphi)}, \quad (\text{D.18e})$$

$$r_2^{(\sigma)} = d_2^{(\sigma)}, \quad r^{(\mathbf{E} \cdot \mathbf{B})} = d^{(\mathbf{E} \cdot \mathbf{B})}. \quad (\text{D.18f})$$

Particular case: No backreaction

In Secs. 8.2 and 8.3, we study the particular scenario where gauge-field backreaction terms are explicitly set to zero in Eq. (D.17). This means, setting $\langle \mathbf{E}^2 \rangle = \langle \mathbf{B}^2 \rangle = \langle \mathbf{E} \cdot \mathbf{B} \rangle = \langle \mathbf{E} \cdot \text{rot } \mathbf{B} \rangle = 0$ in the above equations. The resulting expressions for the coefficients in Eq. (D.18) are then substantially simplified. For $p_{\mathbf{k}}$ and $q_{\mathbf{k}}$ one obtains the simple relationships given in Eq. (8.20), while for the $r^{(\alpha)}$'s, one gets

$$r^{(q)} = \frac{4M_{\text{P}}^2 \mathcal{H}}{\varphi'^2} \left[\frac{k^2}{3} + \frac{a^2 V_{,\phi}(\varphi)}{\varphi'} \left(2\mathcal{H} - \frac{\mathcal{H}'}{\mathcal{H}} - \frac{\varphi''}{\varphi'} + \frac{V_{,\phi\phi}(\varphi)\varphi'}{V_{,\phi}(\varphi)} \right) \right], \quad (\text{D.19a})$$

$$r^{(\rho)} = \frac{4M_{\text{P}}^2 \mathcal{H}}{3\varphi'^2} \left(-2\mathcal{H} + \frac{\mathcal{H}'}{\mathcal{H}} - \frac{\varphi''}{\varphi'} \right), \quad (\text{D.19b})$$

$$r_1^{(\sigma)} = \frac{4M_{\text{P}}^2 k^2 \mathcal{H}}{3\varphi'^2} \left(-\mathcal{H} + \frac{\mathcal{H}'}{\mathcal{H}} + 2\frac{\varphi''}{\varphi'} \right), \quad (\text{D.19c})$$

$$r_2^{(\sigma)} = -\frac{4M_{\text{P}}^2 k^2 \mathcal{H}}{3\varphi'^2}, \quad (\text{D.19d})$$

$$r^{(\mathbf{E} \cdot \mathbf{B})} = \frac{2}{3} a^2 \frac{\beta}{M_{\text{P}}} \frac{\mathcal{H}}{\varphi'}. \quad (\text{D.19e})$$

D.2 Expressions for the source functions K and L

The source term on the right-hand side of Eq. (D.17) is conveniently decomposed into five contributions. Furthermore, in the main text, this source term was further expanded in terms of gauge-field

mode functions and creation and annihilation operators, see Eq. (8.35). The functions K and L appearing in this equation may themselves also be decomposed into five terms, each proportional to one of the coefficients $r^{(\alpha)}$ in Eq. (D.17). These terms are given by

$$K_{\lambda, \mathbf{p}; \lambda', \mathbf{p}'; \mathbf{k}}^{(1)}(\eta) = \frac{a^2 r^{(q)}}{4M_{\text{P}}^2} \frac{i}{k^2} (\mathbf{k} \cdot [\boldsymbol{\epsilon}_{\lambda}(\mathbf{p}) \times \boldsymbol{\epsilon}_{\lambda'}(\mathbf{p}')]) \left[\mathcal{E}_{\lambda}(\eta, p) \mathcal{B}_{\lambda'}(\eta, p') - \mathcal{B}_{\lambda}(\eta, p) \mathcal{E}_{\lambda'}(\eta, p') \right], \quad (\text{D.20a})$$

$$K_{\lambda, \mathbf{p}; \lambda', \mathbf{p}'; \mathbf{k}}^{(2)}(\eta) = \frac{a^2 r^{(\rho)}}{4M_{\text{P}}^2} (\boldsymbol{\epsilon}_{\lambda}(\mathbf{p}) \cdot \boldsymbol{\epsilon}_{\lambda'}(\mathbf{p}')) \left[\mathcal{E}_{\lambda}(\eta, p) \mathcal{E}_{\lambda'}(\eta, p') + \mathcal{B}_{\lambda}(\eta, p) \mathcal{B}_{\lambda'}(\eta, p') \right], \quad (\text{D.20b})$$

$$K_{\lambda, \mathbf{p}; \lambda', \mathbf{p}'; \mathbf{k}}^{(3)}(\eta) = -\frac{a^2 r_1^{(\sigma)}}{4M_{\text{P}}^2 k^2} \left(\delta_{ij} - 3 \frac{k_i k_j}{k^2} \right) \epsilon_{\lambda}^i(\mathbf{p}) \epsilon_{\lambda'}^j(\mathbf{p}') \times \left[\mathcal{E}_{\lambda}(\eta, p) \mathcal{E}_{\lambda'}(\eta, p') + \mathcal{B}_{\lambda}(\eta, p) \mathcal{B}_{\lambda'}(\eta, p') \right], \quad (\text{D.20c})$$

$$K_{\lambda, \mathbf{p}; \lambda', \mathbf{p}'; \mathbf{k}}^{(4)}(\eta) = \frac{a^2 r_2^{(\sigma)}}{4M_{\text{P}}^2 k^2} \left(\delta_{ij} - 3 \frac{k_i k_j}{k^2} \right) \epsilon_{\lambda}^i(\mathbf{p}) \epsilon_{\lambda'}^j(\mathbf{p}') \times \left\{ \left(2\mathcal{H} + \frac{\beta}{M_{\text{P}}} \varphi' \right) \left[\mathcal{E}_{\lambda}(\eta, p) \mathcal{B}_{\lambda'}(\eta, p') + \mathcal{B}_{\lambda}(\eta, p) \mathcal{E}_{\lambda'}(\eta, p') \right] + (\lambda p - \lambda' p') \left[\mathcal{E}_{\lambda}(\eta, p) \mathcal{B}_{\lambda'}(\eta, p') - \mathcal{B}_{\lambda}(\eta, p) \mathcal{E}_{\lambda'}(\eta, p') \right] \right\}, \quad (\text{D.20d})$$

$$K_{\lambda, \mathbf{p}; \lambda', \mathbf{p}'; \mathbf{k}}^{(5)}(\eta) = \frac{r^{(\mathbf{E} \cdot \mathbf{B})}}{2} (\boldsymbol{\epsilon}_{\lambda}(\mathbf{p}) \cdot \boldsymbol{\epsilon}_{\lambda'}(\mathbf{p}')) \left[\mathcal{E}_{\lambda}(\eta, p) \mathcal{B}_{\lambda'}(\eta, p') + \mathcal{B}_{\lambda}(\eta, p) \mathcal{E}_{\lambda'}(\eta, p') \right]. \quad (\text{D.20e})$$

Entirely analogous expressions can be found for the source function L :

$$L_{\lambda, \mathbf{p}; \lambda', \mathbf{p}'; \mathbf{k}}^{(1)}(\eta) = \frac{a^2 r^{(q)}}{4M_{\text{P}}^2} \frac{i}{k^2} (\mathbf{k} \cdot [\boldsymbol{\epsilon}_{\lambda}^*(\mathbf{p}) \times \boldsymbol{\epsilon}_{\lambda'}(\mathbf{p}')]) \left[\mathcal{E}_{\lambda}^*(\eta, p) \mathcal{B}_{\lambda'}(\eta, p') - \mathcal{B}_{\lambda}^*(\eta, p) \mathcal{E}_{\lambda'}(\eta, p') \right], \quad (\text{D.21a})$$

$$L_{\lambda, \mathbf{p}; \lambda', \mathbf{p}'; \mathbf{k}}^{(2)}(\eta) = \frac{a^2 r^{(\rho)}}{4M_{\text{P}}^2} (\boldsymbol{\epsilon}_{\lambda}^*(\mathbf{p}) \cdot \boldsymbol{\epsilon}_{\lambda'}(\mathbf{p}')) \left[\mathcal{E}_{\lambda}^*(\eta, p) \mathcal{E}_{\lambda'}(\eta, p') + \mathcal{B}_{\lambda}^*(\eta, p) \mathcal{B}_{\lambda'}(\eta, p') \right], \quad (\text{D.21b})$$

$$L_{\lambda, \mathbf{p}; \lambda', \mathbf{p}'; \mathbf{k}}^{(3)}(\eta) = -\frac{a^2 r_1^{(\sigma)}}{4M_{\text{P}}^2 k^2} \left(\delta_{ij} - 3 \frac{k_i k_j}{k^2} \right) \epsilon_{\lambda}^{*i}(\mathbf{p}) \epsilon_{\lambda'}^j(\mathbf{p}') \times \left[\mathcal{E}_{\lambda}^*(\eta, p) \mathcal{E}_{\lambda'}(\eta, p') + \mathcal{B}_{\lambda}^*(\eta, p) \mathcal{B}_{\lambda'}(\eta, p') \right], \quad (\text{D.21c})$$

$$L_{\lambda, \mathbf{p}; \lambda', \mathbf{p}'; \mathbf{k}}^{(4)}(\eta) = \frac{a^2 r_2^{(\sigma)}}{4M_{\text{P}}^2 k^2} \left(\delta_{ij} - 3 \frac{k_i k_j}{k^2} \right) \epsilon_{\lambda}^{*i}(\mathbf{p}) \epsilon_{\lambda'}^j(\mathbf{p}') \times \left\{ \left(2\mathcal{H} + \frac{\beta}{M_{\text{P}}} \varphi' \right) \left[\mathcal{E}_{\lambda}^*(\eta, p) \mathcal{B}_{\lambda'}(\eta, p') + \mathcal{B}_{\lambda}^*(\eta, p) \mathcal{E}_{\lambda'}(\eta, p') \right] + (\lambda p - \lambda' p') \left[\mathcal{E}_{\lambda}^*(\eta, p) \mathcal{B}_{\lambda'}(\eta, p') - \mathcal{B}_{\lambda}^*(\eta, p) \mathcal{E}_{\lambda'}(\eta, p') \right] \right\}, \quad (\text{D.21d})$$

$$L_{\lambda, \mathbf{p}; \lambda', \mathbf{p}'; \mathbf{k}}^{(5)}(\eta) = \frac{r^{(\mathbf{E} \cdot \mathbf{B})}}{2} (\boldsymbol{\epsilon}_{\lambda}^*(\mathbf{p}) \cdot \boldsymbol{\epsilon}_{\lambda'}(\mathbf{p}')) \left[\mathcal{E}_{\lambda}^*(\eta, p) \mathcal{B}_{\lambda'}(\eta, p') + \mathcal{B}_{\lambda}^*(\eta, p) \mathcal{E}_{\lambda'}(\eta, p') \right]. \quad (\text{D.21e})$$

The final source terms are then

$$K_{\lambda, \mathbf{p}; \lambda', \mathbf{p}'; \mathbf{k}}(\eta) = \sum_{N=1}^5 K_{\lambda, \mathbf{p}; \lambda', \mathbf{p}'; \mathbf{k}}^{(N)}(\eta), \quad L_{\lambda, \mathbf{p}; \lambda', \mathbf{p}'; \mathbf{k}}(\eta) = \sum_{N=1}^5 L_{\lambda, \mathbf{p}; \lambda', \mathbf{p}'; \mathbf{k}}^{(N)}(\eta) \quad (\text{D.22})$$

In our notation, the source terms used in Ref. [57] are given by

$$\begin{aligned} \tilde{K}_{\lambda,p;\lambda',p';k}(\eta) = \frac{1}{2} \left(r^{(\mathbf{E} \cdot \mathbf{B})} - \frac{\beta}{M_{\text{P}}} \frac{a^2 r_2^{(\sigma)} \varphi'}{4M_{\text{P}}^2 k^2} \right) (\epsilon_{\lambda}(\mathbf{p}) \cdot \epsilon_{\lambda'}(\mathbf{p}')) \\ \times \left[\mathcal{E}_{\lambda}(\eta, p) \mathcal{B}_{\lambda'}(\eta, p') + \mathcal{B}_{\lambda}(\eta, p) \mathcal{E}_{\lambda'}(\eta, p') \right], \end{aligned} \quad (\text{D.23a})$$

$$\begin{aligned} \tilde{L}_{\lambda,p;\lambda',p';k}(\eta) = \frac{1}{2} \left(r^{(\mathbf{E} \cdot \mathbf{B})} - \frac{\beta}{M_{\text{P}}} \frac{a^2 r_2^{(\sigma)} \varphi'}{4M_{\text{P}}^2 k^2} \right) (\epsilon_{\lambda}(\mathbf{p}) \cdot \epsilon_{\lambda'}(\mathbf{p}')) \\ \times \left[\mathcal{E}_{\lambda}^*(\eta, p) \mathcal{B}_{\lambda'}(\eta, p') + \mathcal{B}_{\lambda}^*(\eta, p) \mathcal{E}_{\lambda'}(\eta, p') \right], \end{aligned} \quad (\text{D.23b})$$

with $r^{(\mathbf{E} \cdot \mathbf{B})}$ and $r_2^{(\sigma)}$ given by the expressions for $\langle \mathbf{E}^2 \rangle = \langle \mathbf{B}^2 \rangle = \langle \mathbf{E} \cdot \mathbf{B} \rangle = 0$, i.e., as given in Eq. (D.19) (see also the discussion below Eq. (8.64)).

Polarization vectors

Let us also discuss the treatment of the polarization vectors appearing in Eqs. (D.20) and (D.21). In chapter 8, the source terms K and L appear in the integral expressions for the induced scalar power spectrum, Eq. (8.39b), and bispectrum, Eq. (8.42), for the case where the momenta $\mathbf{p}$, $\mathbf{p}'$ and $\mathbf{k}$ in Eqs. (D.20) and (D.21) obey $\mathbf{p} + \mathbf{p}' = \mathbf{k}$. As detailed in Ref. [2], this particular momentum configuration allows one to evaluate the polarization vector combinations appearing in Eqs. (D.20) and (D.21). Using the identities for the polarization vectors $\epsilon_{\lambda}(\mathbf{p})$ in Eq. (3.16), one may determine that all expressions involving polarization vectors in Eqs. (D.20) and (D.21) are related as follows

$$\delta_{ij} \epsilon_{\lambda}^i(\mathbf{p}) \epsilon_{\lambda'}^j(\mathbf{p}') = (\epsilon_{\lambda}(\mathbf{p}) \cdot \epsilon_{\lambda'}(\mathbf{p}')), \quad (\text{D.24a})$$

$$k_i k_j \epsilon_{\lambda}^i(\mathbf{p}) \epsilon_{\lambda'}^j(\mathbf{p}') = (\mathbf{p} \cdot \mathbf{p}' + \lambda \lambda' p p') (\epsilon_{\lambda}(\mathbf{p}) \cdot \epsilon_{\lambda'}(\mathbf{p}')), \quad (\text{D.24b})$$

$$i(\mathbf{k} \cdot [\epsilon_{\lambda}(\mathbf{p}) \times \epsilon_{\lambda'}(\mathbf{p}')]) = (\lambda p - \lambda' p') (\epsilon_{\lambda}(\mathbf{p}) \cdot \epsilon_{\lambda'}(\mathbf{p}')). \quad (\text{D.24c})$$

For the particular case of the scalar power spectrum, where two momenta $\mathbf{p}$ and $\mathbf{p}'$ are fully constrained via momentum conservation, one may then derive that [2]

$$(\epsilon_{\lambda}(\mathbf{p}) \cdot \epsilon_{\lambda'}(\mathbf{p}')) = \frac{1}{2} \left(1 - \lambda \lambda' \frac{\mathbf{p} \cdot \mathbf{p}'}{p p'} \right). \quad (\text{D.25})$$

For the calculation of the bispectrum, this relationship is however no longer valid, as there are in total three momenta entering the integral in Eq. 8.42. In this case, one may derive a similar expression for $\epsilon_{\lambda}(\mathbf{p}) \cdot \epsilon_{\lambda'}(\mathbf{p}')$ as detailed in App. E of Ref. [2].

Bibliography

- [1] Richard von Eckardstein et al. “Axion Inflation in the Strong-Backreaction Regime: Decay of the Anber-Sorbo Solution”. In: *JHEP* 11.11 (Nov. 2023), page 183. doi: 10.1007/JHEP11(2023)183. arXiv: 2309.04254 [hep-ph].
- [2] R. Durrer et al. “Scalar Perturbations from Inflation in the Presence of Gauge Fields”. In: *Phys. Rev. D* 110.4 (Aug. 2024), page 043533. doi: 10.1103/PhysRevD.110.043533. arXiv: 2404.19694 [astro-ph.CO].
- [3] Richard von Eckardstein, Kai Schmitz, and Oleksandr Sobol. “On the Schwinger Effect during Axion Inflation”. In: *JHEP* 02.2 (Feb. 2025), page 096. doi: 10.1007/JHEP02(2025)096. arXiv: 2408.16538 [hep-ph].
- [4] Richard von Eckardstein, Kai Schmitz, and Oleksandr Sobol. “Gravitational Waves from Axion Inflation in the Gradient Expansion Formalism I: Pure Axion Inflation”. In: (Aug. 2025). doi: 10.48550/arXiv.2508.00798. arXiv: 2508.00798 [astro-ph.CO].
- [5] Richard von Eckardstein, Kai Schmitz, and Oleksandr Sobol. “Gravitational Waves from Axion Inflation in the Gradient Expansion Formalism II: Fermionic Axion Inflation”. In: (Sept. 2025). doi: 10.48550/arXiv.2509.25013. arXiv: 2509.25013 [astro-ph.CO].
- [6] Oleksandr Sobol et al. “Nonlinear Physics of Axion Inflation”. In: (Mar. 2026). doi: 10.48550/arXiv.2603.02570. arXiv: 2603.02570 [hep-ph].
- [7] Richard von Eckardstein. *GEFF: The Gradient Expansion Formalism Factory*. Oct. 2025. doi: 10.5281/zenodo.17356579. URL: <https://doi.org/10.5281/zenodo.17356579>.
- [8] Gilles Weymann-Despres et al. “MSSM-inflation Revisited: Towards a Coherent Description of High-Energy Physics and Cosmology”. In: *Physical Review D* 108.2 (July 2023), page 023511. doi: 10.1103/PhysRevD.108.023511. arXiv: 2304.04534 [hep-ph].
- [9] Adeela Afzal et al. “The NANOGrav 15 Yr Data Set: Search for Signals from New Physics”. In: *The Astrophysical Journal Letters* 951.1 (July 2023), page L11. doi: 10.3847/2041-8213/acdc91.
- [10] Gabriella Agazie et al. “The NANOGrav 15 Yr Data Set: Running of the Spectral Index”. In: *The Astrophysical Journal Letters* 978.2 (Jan. 2025), page L29. doi: 10.3847/2041-8213/ad99d3.
- [11] Andrea Mitridate et al. “PTArcade”. In: (June 2023). doi: 10.48550/arXiv.2306.16377. arXiv: 2306.16377 [hep-ph].
- [12] Planck Collaboration et al. “Planck 2018 Results. I. Overview and the Cosmological Legacy of Planck”. In: *Astronomy & Astrophysics* 641 (Sept. 2020), A1. doi: 10.1051/0004-6361/201833880. arXiv: 1807.06205 [astro-ph].
- [13] G. F. Smoot et al. “Structure in the COBE Differential Microwave Radiometer First-Year Maps”. In: *The Astrophysical Journal* 396 (1992), pages L1–L5. doi: 10.1086/186504.

- [14] Planck Collaboration et al. “Planck 2018 Results. X. Constraints on Inflation”. In: *Astronomy & Astrophysics* 641 (Sept. 2020), A10. doi: 10.1051/0004-6361/201833887. arXiv: 1807.06211 [astro-ph.CO].
- [15] Planck Collaboration et al. “Planck 2018 Results. IX. Constraints on Primordial Non-Gaussianity”. In: 641 (Sept. 2020), A9. doi: 10.1051/0004-6361/201935891. arXiv: 1905.05697 [astro-ph.CO].
- [16] Planck Collaboration et al. “Planck 2018 Results. VI. Cosmological Parameters”. In: *Astron. Astrophys.* 641 (Sept. 2020), A6. doi: 10.1051/0004-6361/201833910. arXiv: 1807.06209 [astro-ph.CO].
- [17] A.A. Starobinsky. “A New Type of Isotropic Cosmological Models without Singularity”. In: *Phys. Lett. B* 91.1 (1980), pages 99–102. doi: 10.1016/0370-2693(80)90670-X.
- [18] Alan H. Guth. “Inflationary Universe: A Possible Solution to the Horizon and Flatness Problems”. In: *Phys. Rev. D* 23.2 (1981), pages 347–356. doi: 10.1103/PhysRevD.23.347.
- [19] A.D. Linde. “A New Inflationary Universe Scenario: A Possible Solution of the Horizon, Flatness, Homogeneity, Isotropy and Primordial Monopole Problems”. In: *Phys. Lett. B* 108.6 (1982), pages 389–393. doi: 10.1016/0370-2693(82)91219-9.
- [20] Andreas Albrecht and Paul J. Steinhardt. “Cosmology for Grand Unified Theories with Radiatively Induced Symmetry Breaking”. In: *Phys. Rev. Lett.* 48.17 (1982), pages 1220–1223. doi: 10.1103/PhysRevLett.48.1220.
- [21] A.D. Linde. “Chaotic Inflation”. In: *Phys. Lett. B* 129.3-4 (1983), pages 177–181. doi: 10.1016/0370-2693(83)90837-7.
- [22] Viatcheslav F. Mukhanov and G. V. Chibisov. “Quantum Fluctuations and a Nonsingular Universe”. In: *JETP Lett.* 33 (1981), pages 532–535.
- [23] Viatcheslav F. Mukhanov and G. V. Chibisov. “The Vacuum energy and large scale structure of the universe”. In: *Sov. Phys. JETP* 56 (1982), pages 258–265.
- [24] A.A. Starobinsky. “Dynamics of Phase Transition in the New Inflationary Universe Scenario and Generation of Perturbations”. In: *Phys. Lett. B* 117.3-4 (1982), pages 175–178. doi: 10.1016/0370-2693(82)90541-X.
- [25] S.W. Hawking. “The Development of Irregularities in a Single Bubble Inflationary Universe”. In: *Phys. Lett. B* 115.4 (1982), page 295. doi: 10.1016/0370-2693(82)90373-2.
- [26] James M. Bardeen, Paul J. Steinhardt, and Michael S. Turner. “Spontaneous Creation of Almost Scale-Free Density Perturbations in an Inflationary Universe”. In: *Phys. Rev. D* 28.4 (1983), page 679. doi: 10.1103/PhysRevD.28.679.
- [27] L. P. Grishchuk. “Amplification of gravitational waves in an isotropic universe”. In: *Sov. Phys. JETP* 40.3 (1975), pages 409–415.
- [28] Alexei A. Starobinsky. “Spectrum of relict gravitational radiation and the early state of the universe”. In: *JETP Lett.* 30 (1979). Edited by I. M. Khalatnikov and V. P. Mineev, pages 682–685.
- [29] V.A. Rubakov, M.V. Sazhin, and A.V. Veryaskin. “Graviton Creation in the Inflationary Universe and the Grand Unification Scale”. In: *Phys. Lett. B* 115.3 (1982), pages 189–192. doi: 10.1016/0370-2693(82)90641-4.

- [30] R. Fabbri and M.D. Pollock. “The Effect of Primordially Produced Gravitons upon the Anisotropy of the Cosmological Microwave Background Radiation”. In: *Phys. Lett. B* 125.6 (1983), pages 445–448. doi: 10.1016/0370-2693(83)91322-9.
- [31] L.F. Abbott and Mark B. Wise. “Constraints on Generalized Inflationary Cosmologies”. In: *Nucl. Phys. B* 244.2 (1984), pages 541–548. doi: 10.1016/0550-3213(84)90329-8.
- [32] E. Camphuis et al. “SPT-3G D1: CMB Temperature and Polarization Power Spectra and Cosmology from 2019 and 2020 Observations of the SPT-3G Main Field”. In: (June 2025). doi: 10.48550/arXiv.2506.20707. arXiv: 2506.20707 [astro-ph.CO].
- [33] Erminia Calabrese et al. “The Atacama Cosmology Telescope: DR6 Constraints on Extended Cosmological Models”. In: (June 2025). doi: 10.48550/arXiv.2503.14454. arXiv: 2503.14454 [astro-ph.CO].
- [34] Lev Kofman et al. “Beauty Is Attractive: Moduli Trapping at Enhanced Symmetry Points”. In: *Journal of High Energy Physics* 05.05 (2004), page 030. doi: 10.1088/1126-6708/2004/05/030. arXiv: hep-th/0403001.
- [35] Daniel Green et al. “Trapped Inflation”. In: *Physical Review D* 80.6 (2009), page 063533. doi: 10.1103/PhysRevD.80.063533. arXiv: 0902.1006 [hep-th].
- [36] Daniel J. H. Chung et al. “Probing Planckian Physics: Resonant Production of Particles during Inflation and Features in the Primordial Power Spectrum”. In: *Physical Review D* 62.4 (2000), page 043508. doi: 10.1103/PhysRevD.62.043508. arXiv: hep-ph/9910437.
- [37] G. J. Mathews et al. “Constraints on Resonant Particle Production during Inflation from the Matter and CMB Power Spectra”. In: *Physical Review D* 70.8 (2004), page 083505. doi: 10.1103/PhysRevD.70.083505. arXiv: astro-ph/0406046.
- [38] Oystein Elgaroy, Steen Hannestad, and Troels Haugboelle. “Observational Constraints on Particle Production during Inflation”. In: *Journal of Cosmology and Astroparticle Physics* 09.09 (2003), page 008. doi: 10.1088/1475-7516/2003/09/008. arXiv: astro-ph/0306229.
- [39] Antonio Enea Romano and Misao Sasaki. “Effects of Particle Production during Inflation”. In: *Physical Review D* 78.10 (2008), page 103522. doi: 10.1103/PhysRevD.78.103522. arXiv: 0809.5142 [gr-qc].
- [40] David Langlois and Lorenzo Sorbo. “Primordial Perturbations and Non-Gaussianities from Modulated Trapping”. In: *Journal of Cosmology and Astroparticle Physics* 08.08 (2009), page 014. doi: 10.1088/1475-7516/2009/08/014. arXiv: 0906.1813 [astro-ph.CO].
- [41] A D Dolgov and D P Kirilova. “On Particle Creation by a Time Dependent Scalar Field”. In: *Sov. J. Nucl. Phys.* 51 (1990), pages 172–177.
- [42] Jennie H. Traschen and Robert H. Brandenberger. “Particle Production during Out-of-Equilibrium Phase Transitions”. In: *Physical Review D* 42.8 (1990), pages 2491–2504. doi: 10.1103/PhysRevD.42.2491.
- [43] Lev Kofman, Andrei Linde, and Alexei Starobinsky. “Reheating after Inflation”. In: *Physical Review Letters* 73.24 (1994), pages 3195–3198. doi: 10.1103/PhysRevLett.73.3195. arXiv: hep-th/9405187.
- [44] Lev Kofman, Andrei Linde, and Alexei Starobinsky. “Towards the Theory of Reheating After Inflation”. In: *Physical Review D* 56.6 (1997), pages 3258–3295. doi: 10.1103/PhysRevD.56.3258. arXiv: hep-ph/9704452.

- [45] Patrick Greene et al. “Structure of Resonance in Preheating after Inflation”. In: *Physical Review D* 56.10 (1997), pages 6175–6192. doi: 10.1103/PhysRevD.56.6175. arXiv: hep-ph/9705347.
- [46] Neil Barnaby et al. “Cosmological Fluctuations from Infra-Red Cascading During Inflation”. In: *Physical Review D* 80.4 (2009), page 043501. doi: 10.1103/PhysRevD.80.043501. arXiv: 0902.0615 [hep-th].
- [47] Neil Barnaby and Zhiqi Huang. “Particle Production During Inflation: Observational Constraints and Signatures”. In: *Physical Review D* 80.12 (2009), page 126018. doi: 10.1103/PhysRevD.80.126018. arXiv: 0909.0751 [astro-ph.CO].
- [48] Neil Barnaby. “On Features and Nongaussianity from Inflationary Particle Production”. In: *Physical Review D* 82.10 (2010), page 106009. doi: 10.1103/PhysRevD.82.106009. arXiv: 1006.4615 [astro-ph.CO].
- [49] Planck Collaboration et al. “Planck 2015 Results. XVII. Constraints on Primordial Non-Gaussianity”. In: *Astronomy & Astrophysics* 594 (Sept. 2016), A17. doi: 10.1051/0004-6361/201525836. arXiv: 1502.01592 [astro-ph.CO].
- [50] Leonardo Senatore, Eva Silverstein, and Matias Zaldarriaga. “New Sources of Gravitational Waves during Inflation”. In: *Journal of Cosmology and Astroparticle Physics* 08.08 (2014), page 016. doi: 10.1088/1475-7516/2014/08/016. arXiv: 1109.0542 [hep-th].
- [51] Jessica L. Cook and Lorenzo Sorbo. “Particle Production during Inflation and Gravitational Waves Detectable by Ground-Based Interferometers”. In: *Phys. Rev. D* 85.2 (2012), page 023534. doi: 10.1103/PhysRevD.85.023534. arXiv: 1109.0022 [astro-ph.CO].
- [52] Neil Barnaby et al. “Gravity Waves and Non-Gaussian Features from Particle Production in a Sector Gravitationally Coupled to the Inflaton”. In: *Phys. Rev. D* 86.10 (2012), page 103508. doi: 10.1103/PhysRevD.86.103508. arXiv: 1206.6117 [astro-ph.CO].
- [53] W. Daniel Garretson, George B. Field, and Sean M. Carroll. “Primordial Magnetic Fields from Pseudo-Goldstone Bosons”. In: *Phys. Rev. D* 46 (1992), pages 5346–5351. doi: 10.1103/PhysRevD.46.5346. arXiv: hep-ph/9209238.
- [54] Mohamed M. Anber and Lorenzo Sorbo. “N-Flationary Magnetic Fields”. In: *JCAP* 10 (2006), page 018. doi: 10.1088/1475-7516/2006/10/018. arXiv: astro-ph/0606534.
- [55] Mohamed M. Anber and Lorenzo Sorbo. “Naturally Inflating on Steep Potentials through Electromagnetic Dissipation”. In: *Phys. Rev. D* 81.4 (2010), page 043534. doi: 10.1103/PhysRevD.81.043534. arXiv: 0908.4089 [hep-th].
- [56] Neil Barnaby and Marco Peloso. “Large Nongaussianity in Axion Inflation”. In: *Phys. Rev. Lett.* 106.18 (2011), page 181301. doi: 10.1103/PhysRevLett.106.181301. arXiv: 1011.1500 [hep-ph].
- [57] Neil Barnaby, Ryo Namba, and Marco Peloso. “Phenomenology of a Pseudo-Scalar Inflaton: Naturally Large Nongaussianity”. In: *JCAP* 04.04 (2011), page 009. doi: 10.1088/1475-7516/2011/04/009. arXiv: 1102.4333 [astro-ph.CO].
- [58] Lorenzo Sorbo. “Parity Violation in the Cosmic Microwave Background from a Pseudoscalar Inflaton”. In: *JCAP* 06.06 (2011), page 003. doi: 10.1088/1475-7516/2011/06/003. arXiv: 1101.1525 [astro-ph.CO].

- [59] Katherine Freese, Joshua A. Frieman, and Angela V. Olinto. “Natural Inflation with Pseudo Nambu-Goldstone Bosons”. In: *Phys. Rev. Lett.* 65.26 (1990), pages 3233–3236. doi: 10.1103/PhysRevLett.65.3233.
- [60] Jihn E. Kim, Hans Peter Nilles, and Marco Peloso. “Completing Natural Inflation”. In: *Journal of Cosmology and Astroparticle Physics* 01.01 (2005), page 005. doi: 10.1088/1475-7516/2005/01/005. arXiv: hep-ph/0409138.
- [61] Savas Dimopoulos et al. “N-Flation”. In: *Journal of Cosmology and Astroparticle Physics* 08.08 (2008), page 003. doi: 10.1088/1475-7516/2008/08/003. arXiv: hep-th/0507205.
- [62] Eva Silverstein and Alexander Westphal. “Monodromy in the CMB: Gravity Waves and String Inflation”. In: *Physical Review D* 78.10 (2008), page 106003. doi: 10.1103/PhysRevD.78.106003. arXiv: 0803.3085 [hep-th].
- [63] Liam McAllister, Eva Silverstein, and Alexander Westphal. “Gravity Waves and Linear Inflation from Axion Monodromy”. In: *Physical Review D* 82.4 (2010), page 046003. doi: 10.1103/PhysRevD.82.046003. arXiv: 0808.0706 [hep-th].
- [64] Peter Adshead and Mark Wyman. “Chromo-Natural Inflation: Natural Inflation on a Steep Potential with Classical Non-Abelian Gauge Fields”. In: *Phys. Rev. Lett.* 108.26 (2012), page 261302. doi: 10.1103/PhysRevLett.108.261302. arXiv: 1202.2366 [hep-th].
- [65] Enrico Pajer and Marco Peloso. “A Review of Axion Inflation in the Era of Planck”. In: *Classical and Quantum Gravity* 30.21 (Oct. 2013), page 214002. doi: 10.1088/0264-9381/30/21/214002. arXiv: 1305.3557 [hep-th].
- [66] John Preskill, Mark B. Wise, and Frank Wilczek. “Cosmology of the Invisible Axion”. In: *Phys. Lett. B* 120 (1983). Edited by M. A. Srednicki, pages 127–132. doi: 10.1016/0370-2693(83)90637-8.
- [67] L. F. Abbott and P. Sikivie. “A Cosmological Bound on the Invisible Axion”. In: *Phys. Lett. B* 120 (1983). Edited by M. A. Srednicki, pages 133–136. doi: 10.1016/0370-2693(83)90638-X.
- [68] Michael Dine and Willy Fischler. “The Not So Harmless Axion”. In: *Phys. Lett. B* 120 (1983). Edited by M. A. Srednicki, pages 137–141. doi: 10.1016/0370-2693(83)90639-1.
- [69] Steven Weinberg. “A New Light Boson?” In: *Phys. Rev. Lett.* 40 (1978), pages 223–226. doi: 10.1103/PhysRevLett.40.223.
- [70] Frank Wilczek. “Problem of Strong P and T Invariance in the Presence of Instantons”. In: *Phys. Rev. Lett.* 40 (1978), pages 279–282. doi: 10.1103/PhysRevLett.40.279.
- [71] R. D. Peccei and Helen R. Quinn. “CP Conservation in the Presence of Instantons”. In: *Phys. Rev. Lett.* 38 (1977), pages 1440–1443. doi: 10.1103/PhysRevLett.38.1440.
- [72] R. D. Peccei and Helen R. Quinn. “Constraints Imposed by CP Conservation in the Presence of Instantons”. In: *Phys. Rev. D* 16 (1977), pages 1791–1797. doi: 10.1103/PhysRevD.16.1791.
- [73] R. D. Peccei. “The Strong CP problem and axions”. In: *Lect. Notes Phys.* 741 (2008). Edited by Markus Kuster, Georg Raffelt, and Berta Beltran, pages 3–17. doi: 10.1007/978-3-540-73518-2_1. arXiv: hep-ph/0607268.

- [74] Georg G. Raffelt. “Astrophysical axion bounds”. In: *Lect. Notes Phys.* 741 (2008). Edited by Markus Kuster, Georg Raffelt, and Berta Beltran, pages 51–71. doi: 10.1007/978-3-540-73518-2_3. arXiv: hep-ph/0611350.
- [75] Pierre Sikivie. “Axion Cosmology”. In: *Lect. Notes Phys.* 741 (2008). Edited by Markus Kuster, Georg Raffelt, and Berta Beltran, pages 19–50. doi: 10.1007/978-3-540-73518-2_2. arXiv: astro-ph/0610440.
- [76] David J. E. Marsh. “Axion Cosmology”. In: *Phys. Rept.* 643 (2016), pages 1–79. doi: 10.1016/j.physrep.2016.06.005. arXiv: 1510.07633 [astro-ph.CO].
- [77] Luca Di Luzio et al. “The landscape of QCD axion models”. In: *Phys. Rept.* 870 (2020), pages 1–117. doi: 10.1016/j.physrep.2020.06.002. arXiv: 2003.01100 [hep-ph].
- [78] Peter Adshead et al. “Gauge-Preheating and the End of Axion Inflation”. In: *JCAP* 12.12 (Dec. 2015), page 034. doi: 10.1088/1475-7516/2015/12/034. arXiv: 1502.06506 [astro-ph.CO].
- [79] Daniel G. Figueroa and Mikhail Shaposhnikov. “Lattice Implementation of Abelian Gauge Theories with Chern-Simons Number and an Axion Field”. In: *Nucl. Phys. B* 926 (Jan. 2018), pages 544–569. doi: 10.1016/j.nuclphysb.2017.12.001. arXiv: 1705.09629 [hep-lat].
- [80] Peter Adshead, John T. Giblin Jr, and Zachary J. Weiner. “Gravitational Waves from Gauge Preheating”. In: *Phys. Rev. D* 98.4 (Aug. 2018), page 4. doi: 10.1103/PhysRevD.98.043525. arXiv: 1805.04550 [astro-ph.CO].
- [81] Jose Roberto Canivete Cuissa and Daniel G. Figueroa. “Lattice Formulation of Axion Inflation. Application to Preheating”. In: *JCAP* 06.06 (June 2019), page 002. doi: 10.1088/1475-7516/2019/06/002. arXiv: 1812.03132 [astro-ph.CO].
- [82] Valerie Domcke, Mauro Pieroni, and Pierre Binétruy. “Primordial Gravitational Waves for Universality Classes of Pseudoscalar Inflation”. In: *JCAP* 06.06 (June 2016), page 031. doi: 10.1088/1475-7516/2016/06/031. arXiv: 1603.01287 [astro-ph.CO].
- [83] Juan Garcia-Bellido et al. “A Flashing Beacon in Axion Inflation: Recurring Bursts of Gravitational Waves in the Strong Backreaction Regime”. In: *JCAP* 01 (Jan. 2024), page 034. doi: 10.1088/1475-7516/2024/01/034. arXiv: 2303.13425 [astro-ph.CO].
- [84] Juan Garcia-Bellido, Marco Peloso, and Caner Unal. “Gravitational Waves at Interferometer Scales and Primordial Black Holes in Axion Inflation”. In: *JCAP* 12.12 (Dec. 2016), page 031. doi: 10.1088/1475-7516/2016/12/031. arXiv: 1610.03763 [astro-ph.CO].
- [85] Mar Bastero-Gil and António Torres Manso. “Parity Violating Gravitational Waves at the End of Inflation”. In: *JCAP* 08.08 (Aug. 2023), page 001. doi: 10.1088/1475-7516/2023/08/001. arXiv: 2209.15572 [gr-qc].
- [86] Martin Teuscher et al. “Gravitational Waves Sourced by Gauge Fields during Inflation”. In: (Oct. 2025). doi: 10.48550/arXiv.2510.00869. arXiv: 2510.00869 [astro-ph.CO].
- [87] Matteo Barbon, Nadir Ijaz, and Marco Peloso. “Axion Inflation in the Regime of Homogeneous Backreaction”. In: (Oct. 2025). doi: 10.48550/arXiv.2510.17207. arXiv: 2510.17207 [astro-ph.CO].

- [88] Federico Greco. “Production of Gravitational Waves by Inflationary Transitions in Aligned Natural Inflation”. In: (Oct. 2025). doi: 10.48550/arXiv.2510.17556. arXiv: 2510.17556 [astro-ph.CO].
- [89] Andrei Linde, Sander Mooij, and Enrico Pajer. “Gauge Field Production in SUGRA Inflation: Local Non-Gaussianity and Primordial Black Holes”. In: *Phys. Rev. D* 87.10 (May 2013), page 103506. doi: 10.1103/PhysRevD.87.103506. arXiv: 1212.1693 [hep-th].
- [90] Edgar Bugaev and Peter Klimai. “Axion Inflation with Gauge Field Production and Primordial Black Holes”. In: *Phys. Rev. D* 90.10 (Nov. 2014), page 103501. doi: 10.1103/PhysRevD.90.103501. arXiv: 1312.7435 [astro-ph.CO].
- [91] Bernard J. Carr and S. W. Hawking. “Black holes in the early Universe”. In: *Mon. Not. Roy. Astron. Soc.* 168 (1974), pages 399–415. doi: 10.1093/mnras/168.2.399.
- [92] Bernard J. Carr. “The Primordial black hole mass spectrum”. In: *Astrophys. J.* 201 (1975), pages 1–19. doi: 10.1086/153853.
- [93] George F. Chapline. “Cosmological effects of primordial black holes”. In: *Nature* 253.5489 (1975), pages 251–252. doi: 10.1038/253251a0.
- [94] Axel Brandenburg et al. “Magnetogenesis from Axion-SU(2) Inflation”. In: *Journal of Cosmology and Astroparticle Physics* 12.12 (Dec. 2024), page 057. doi: 10.1088/1475-7516/2024/12/057. arXiv: 2408.17413 [astro-ph.CO].
- [95] Andrii Neronov and Ievgen Vovk. “Evidence for Strong Extragalactic Magnetic Fields from Fermi Observations of TeV Blazars”. In: *Science* 328.5974 (2010), pages 73–75. doi: 10.1126/science.1184192. arXiv: 1006.3504 [astro-ph.HE].
- [96] A. M. Taylor, I. Vovk, and A. Neronov. “EGMF Constraints from Simultaneous GeV-TeV Observations of Blazars”. In: *Astronomy & Astrophysics* 529 (2011), A144. doi: 10.1051/0004-6361/201116441. arXiv: 1101.0932 [astro-ph.HE].
- [97] Justin Finke, Luis Reyes, and Markos Georganopoulos. “Constraints on the Intergalactic Magnetic Field from Gamma-Ray Observations of Blazars”. In: C121028 (2012), page 365. doi: 10.48550/arXiv.1303.5093. arXiv: 1303.5093 [astro-ph.HE].
- [98] R. Durrer and A. Neronov. “Cosmological Magnetic Fields: Their Generation, Evolution and Observation”. In: *Astron. Astrophys. Rev.* 21.1 (2013), page 62. doi: 10.1007/s00159-013-0062-7. arXiv: 1303.7121 [astro-ph.CO].
- [99] Kandaswamy Subramanian. “The Origin, Evolution and Signatures of Primordial Magnetic Fields”. In: *Rept. Prog. Phys.* 79.7 (May 2016), page 076901. doi: 10.1088/0034-4885/79/7/076901. arXiv: 1504.02311 [astro-ph.CO].
- [100] Tanmay Vachaspati. “Progress on Cosmological Magnetic Fields”. In: *Rept. Prog. Phys.* 84.7 (June 2021), page 074901. doi: 10.1088/1361-6633/ac03a9. arXiv: 2010.10525 [astro-ph.CO].
- [101] Azadeh Maleknejad, Mahdiyar Noorbala, and M. M. Sheikh-Jabbari. “Leptogenesis in Inflationary Models with Non-Abelian Gauge Fields”. In: *General Relativity and Gravitation* 50.9 (Aug. 2018), page 110. doi: 10.1007/s10714-018-2435-8. arXiv: 1208.2807 [hep-th].
- [102] Mohamed M. Anber and Eray Sabancilar. “Hypermagnetic Fields and Baryon Asymmetry from Pseudoscalar Inflation”. In: *Phys. Rev. D* 92.10 (Nov. 2015), page 101501. doi: 10.1103/PhysRevD.92.101501. arXiv: 1507.00744 [hep-th].

- [103] Daniel Jimenez et al. “Baryon Asymmetry and Gravitational Waves from Pseudoscalar Inflation”. In: *JCAP* 12.12 (Dec. 2017), page 011. doi: 10.1088/1475-7516/2017/12/011. arXiv: 1707.07943 [hep-ph].
- [104] Yann Cado and Mariano Quirós. “Numerical Study of the Schwinger Effect in Axion Inflation”. In: *Phys. Rev. D* 106.12 (Dec. 2022), page 123527. doi: 10.1103/PhysRevD.106.123527. arXiv: 2208.10977 [hep-ph].
- [105] Valerie Domcke et al. “Wash-in Leptogenesis after Axion Inflation”. In: *JHEP* 01.1 (Jan. 2023), page 053. doi: 10.1007/JHEP01(2023)053. arXiv: 2210.06412 [hep-ph].
- [106] Tomohiro Fujita et al. “Consistent Generation of Magnetic Fields in Axion Inflation Models”. In: *JCAP* 05.05 (May 2015), page 054. doi: 10.1088/1475-7516/2015/05/054. arXiv: 1503.05802 [astro-ph.CO].
- [107] Shu-Lin Cheng, Wolung Lee, and Kin-Wang Ng. “A Numerical Study of Pseudoscalar Inflation with an Axion-Gauge Field Coupling”. In: *Phys. Rev. D* 93.6 (Mar. 2016), page 063510. doi: 10.1103/PhysRevD.93.063510. arXiv: 1508.00251 [astro-ph.CO].
- [108] Alessio Notari and Konrad Tywoniuk. “Dissipative Axial Inflation”. In: *JCAP* 12.12 (Dec. 2016), page 038. doi: 10.1088/1475-7516/2016/12/038. arXiv: 1608.06223 [hep-th].
- [109] Valerie Domcke et al. “Resonant Backreaction in Axion Inflation”. In: *JCAP* 09 (Sept. 2020), page 009. doi: 10.1088/1475-7516/2020/09/009. arXiv: 2002.02952 [astro-ph.CO].
- [110] Fritz Sauter. “Über das Verhalten eines Elektrons im homogenen elektrischen Feld nach der relativistischen Theorie Diracs”. In: *Z. Phys.* 69.11-12 (1931), pages 742–764. doi: 10.1007/BF01339461.
- [111] W. Heisenberg and H. Euler. “Consequences of Dirac Theory of the Positron”. In: *Z. Phys.* 98.11-12 (Nov. 1936), pages 714–732. doi: 10.1007/BF01343663. arXiv: physics/0605038.
- [112] Julian Schwinger. “On Gauge Invariance and Vacuum Polarization”. In: *Phys. Rev.* 82.5 (1951), pages 664–679. doi: 10.1103/PhysRev.82.664.
- [113] Gerald V. Dunne. “Heisenberg-Euler Effective Lagrangians : Basics and Extensions”. In: (2004). doi: 10.1142/9789812775344_0014. arXiv: hep-th/0406216.
- [114] Remo Ruffini, Gregory Vereshchagin, and She-Sheng Xue. “Electron-Positron Pairs in Physics and Astrophysics: From Heavy Nuclei to Black Holes”. In: *Physics Reports* 487.1-4 (2010), pages 1–140. doi: 10.1016/j.physrep.2009.10.004. arXiv: 0910.0974 [astro-ph.HE].
- [115] Francois Gelis and Naoto Tanji. “Schwinger Mechanism Revisited”. In: *Progress in Particle and Nuclear Physics* 87 (Mar. 2016), pages 1–49. doi: 10.1016/j.pnpnp.2015.11.001. arXiv: 1510.05451 [hep-ph].
- [116] Markus B. Fröb et al. “Schwinger Effect in de Sitter Space”. In: *JCAP* 04.04 (Apr. 2014), page 009. doi: 10.1088/1475-7516/2014/04/009. arXiv: 1401.4137 [hep-th].
- [117] Takeshi Kobayashi and Niayesh Afshordi. “Schwinger Effect in 4D de Sitter Space and Constraints on Magnetogenesis in the Early Universe”. In: *JHEP* 10.10 (Oct. 2014), page 166. doi: 10.1007/JHEP10(2014)166. arXiv: 1408.4141 [hep-th].
- [118] Clément Stahl, Eckhard Strobel, and She-Sheng Xue. “Fermionic Current and Schwinger Effect in de Sitter Spacetime”. In: *Phys. Rev. D* 93.2 (Jan. 2016), page 025004. doi: 10.1103/PhysRevD.93.025004. arXiv: 1507.01686 [gr-qc].

- [119] Ehsan Bavarsad et al. “Effect of a Magnetic Field on Schwinger Mechanism in de Sitter Spacetime”. In: *Phys. Rev. D* 97.2 (Jan. 2018), page 025017. doi: 10.1103/PhysRevD.97.025017. arXiv: 1707.03975 [hep-th].
- [120] Valerie Domcke and Kyohei Mukaida. “Gauge Field and Fermion Production during Axion Inflation”. In: *JCAP* 11.11 (Nov. 2018), page 020. doi: 10.1088/1475-7516/2018/11/020. arXiv: 1806.08769 [hep-ph].
- [121] Valerie Domcke, Yohei Ema, and Kyohei Mukaida. “Chiral Anomaly, Schwinger Effect, Euler-Heisenberg Lagrangian, and Application to Axion Inflation”. In: *JHEP* 02.2 (Feb. 2020), page 055. doi: 10.1007/JHEP02(2020)055. arXiv: 1910.01205 [hep-ph].
- [122] Kari Enqvist, Daniel G. Figueroa, and Tuukka Meriniemi. “Stochastic Background of Gravitational Waves from Fermions”. In: *Phys. Rev. D* 86.6 (2012), page 061301. doi: 10.1103/PhysRevD.86.061301. arXiv: 1203.4943 [astro-ph.CO].
- [123] Daniel G. Figueroa and Tuukka Meriniemi. “Stochastic Background of Gravitational Waves from Fermions – Theory and Applications”. In: *JHEP* 10.10 (2013), page 101. doi: 10.1007/JHEP10(2013)101. arXiv: 1306.6911 [astro-ph.CO].
- [124] Walter Tangarife et al. “Dynamics of Relaxed Inflation”. In: *Journal of High Energy Physics* 02.2 (Feb. 2018), page 084. doi: 10.1007/JHEP02(2018)084. arXiv: 1706.03072 [hep-ph].
- [125] Mar Bastero-Gil et al. “Schwinger Dark Matter Production”. In: *JCAP* 10 (Oct. 2024), page 078. doi: 10.1088/1475-7516/2024/10/078. arXiv: 2312.15137 [hep-ph].
- [126] Mar Bastero-Gil et al. “Super Heavy Dark Matter from Inflationary Schwinger Production”. In: *Phys. Rev. D* 110.9 (Nov. 2024), page 095019. doi: 10.1103/PhysRevD.110.095019. arXiv: 2311.09475 [hep-ph].
- [127] O. O. Sobol, E. V. Gorbar, and S. I. Vilchinskii. “Backreaction of Electromagnetic Fields and the Schwinger Effect in Pseudoscalar Inflation Magnetogenesis”. In: *Phys. Rev. D* 100.6 (Sept. 2019), page 063523. doi: 10.1103/PhysRevD.100.063523. arXiv: 1907.10443 [astro-ph.CO].
- [128] E. V. Gorbar et al. “Gauge-Field Production during Axion Inflation in the Gradient Expansion Formalism”. In: *Phys. Rev. D* 104.12 (Dec. 2021), page 123504. doi: 10.1103/PhysRevD.104.123504. arXiv: 2109.01651 [hep-ph].
- [129] Tomohiro Fujita et al. “Effective Treatment of U(1) Gauge Field and Charged Particles in Axion Inflation”. In: *JCAP* 09.09 (Sept. 2022), page 023. doi: 10.1088/1475-7516/2022/09/023. arXiv: 2204.01180 [hep-ph].
- [130] E. V. Gorbar et al. “Hypermagnetogenesis from Axion Inflation: Model-independent Estimates”. In: *Phys. Rev. D* 105.4 (Feb. 2022), page 043530. doi: 10.1103/PhysRevD.105.043530. arXiv: 2111.04712 [hep-ph].
- [131] Marco Peloso and Lorenzo Sorbo. “Instability in Axion Inflation with Strong Backreaction from Gauge Modes”. In: *JCAP* 01.01 (Jan. 2023), page 038. doi: 10.1088/1475-7516/2023/01/038. arXiv: 2209.08131 [astro-ph.CO].
- [132] Valerie Domcke et al. “Baryogenesis from Axion Inflation”. In: *Journal of Cosmology and Astroparticle Physics* 10.10 (Oct. 2019), page 032. doi: 10.1088/1475-7516/2019/10/032. arXiv: 1905.13318 [hep-ph].
- [133] E. V. Gorbar et al. “Chirality Production during Axion Inflation”. In: *Ukr. J. Phys.* 68.11 (Dec. 2023), page 717. doi: 10.15407/ujpe68.11.717. arXiv: 2111.05848 [hep-ph].

- [134] Oksana Iarygina, Evangelos I. Sfakianakis, and Axel Brandenburg. “Schwinger Effect in Axion Inflation on a Lattice”. In: (2025). doi: 10.48550/ARXIV.2506.20538. arXiv: 2506.20538 [astro-ph.CO].
- [135] O. O. Sobol et al. “Gradient Expansion Formalism for Magnetogenesis in the Kinetic Coupling Model”. In: *Phys. Rev. D* 102.12 (Dec. 2020), page 123512. doi: 10.1103/PhysRevD.102.123512. arXiv: 2010.13587 [astro-ph.CO].
- [136] Viatcheslav Mukhanov. *Physical Foundations of Cosmology*. 1st edition. Cambridge University Press, 2005. doi: 10.1017/CB09780511790553.
- [137] Dmitry S Gorbunov and Valery A Rubakov. *Introduction to the Theory of the Early Universe: Cosmological Perturbations and Inflationary Theory*. World Scientific Publishing Company, Feb. 2011. doi: 10.1142/7873.
- [138] Daniel Baumann. *Cosmology*. 1st edition. Cambridge University Press, July 2022. doi: 10.1017/9781108937092.
- [139] D. J. Fixsen. “The Temperature of the Cosmic Microwave Background”. In: *The Astrophysical Journal* 707.2 (2009), pages 916–920. doi: 10.1088/0004-637X/707/2/916. arXiv: 0911.1955 [astro-ph.CO].
- [140] 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. arXiv: 1807.06209 [astro-ph.CO].
- [141] T. W. B. Kibble. “Topology of Cosmic Domains and Strings”. In: *J. Phys. A* 9 (1976), pages 1387–1398. doi: 10.1088/0305-4470/9/8/029.
- [142] T. W. B. Kibble. “Some Implications of a Cosmological Phase Transition”. In: *Phys. Rept.* 67 (1980), page 183. doi: 10.1016/0370-1573(80)90091-5.
- [143] W. H. Zurek. “Cosmological Experiments in Superfluid Helium?” In: *Nature* 317 (1985), pages 505–508. doi: 10.1038/317505a0.
- [144] A. Vilenkin and E. P. S. Shellard. *Cosmic Strings and Other Topological Defects*. Cambridge University Press, July 2000.
- [145] S. W. Hawking, I. G. Moss, and J. M. Stewart. “Bubble Collisions in the Very Early Universe”. In: *Physical Review D* 26.10 (1982), page 2681. doi: 10.1103/PhysRevD.26.2681.
- [146] Alan H. Guth and Erick J. Weinberg. “Could the Universe Have Recovered from a Slow First-Order Phase Transition?” In: *Nuclear Physics B* 212.2 (1983), pages 321–364. doi: 10.1016/0550-3213(83)90307-3.
- [147] Paul J. Steinhardt and Michael S. Turner. “Prescription for Successful New Inflation”. In: *Physical Review D* 29.10 (1984), pages 2162–2171. doi: 10.1103/PhysRevD.29.2162.
- [148] D. S. Salopek and J. R. Bond. “Nonlinear Evolution of Long-Wavelength Metric Fluctuations in Inflationary Models”. In: *Physical Review D* 42.12 (1990), pages 3936–3962. doi: 10.1103/PhysRevD.42.3936.
- [149] Andrew R. Liddle, Paul Parsons, and John D. Barrow. “Formalising the Slow-Roll Approximation in Inflation”. In: *Phys. Rev. D* 50.12 (1994), pages 7222–7232. doi: 10.1103/PhysRevD.50.7222. arXiv: astro-ph/9408015.
- [150] Jerome Martin, Christophe Ringeval, and Vincent Vennin. “Encyclopaedia Inflationaris”. In: *Physics of the Dark Universe* 5–6 (Dec. 2014), pages 75–235. doi: 10.1016/j.dark.2014.01.003. doi: 10.1016/j.dark.2024.101653. arXiv: 1303.3787 [astro-ph].

- [151] Lotfi Boubekeur and David H. Lyth. “Hilltop Inflation”. In: *Journal of Cosmology and Astroparticle Physics* 07.07 (2005), page 010. doi: 10.1088/1475-7516/2005/07/010. arXiv: hep-ph/0502047.
- [152] Andrei Linde. “Hybrid Inflation”. In: *Physical Review D* 49.2 (1994), pages 748–754. doi: 10.1103/PhysRevD.49.748. arXiv: astro-ph/9307002.
- [153] David Wands. “Multiple Field Inflation”. In: *Lect. Notes Phys.* Volume 738. 2008, pages 275–304. doi: 10.1007/978-3-540-74353-8_8. arXiv: astro-ph/0702187.
- [154] Andrew R. Liddle, Anupam Mazumdar, and Franz E. Schunck. “Assisted Inflation”. In: *Physical Review D* 58.6 (1998), page 061301. doi: 10.1103/PhysRevD.58.061301. arXiv: astro-ph/9804177.
- [155] Christopher Gordon et al. “Adiabatic and Entropy Perturbations from Inflation”. In: *Physical Review D* 63.2 (Dec. 2000), page 023506. doi: 10.1103/PhysRevD.63.023506. arXiv: astro-ph/0009131.
- [156] Kari Enqvist and Martin S. Sloth. “Adiabatic CMB Perturbations in Pre-Big Bang String Cosmology”. In: *Nuclear Physics B* 626.1-2 (2002), pages 395–409. doi: 10.1016/S0550-3213(02)00043-3. arXiv: hep-ph/0109214.
- [157] David H. Lyth and David Wands. “Generating the Curvature Perturbation without an Inflaton”. In: *Physics Letters B* 524.1-2 (2002), pages 5–14. doi: 10.1016/S0370-2693(01)01366-1. arXiv: hep-ph/0110002.
- [158] Takeo Moroi and Tomo Takahashi. “Effects of Cosmological Moduli Fields on Cosmic Microwave Background”. In: *Physics Letters B* 522.3-4 (2001), pages 215–221. doi: 10.1016/S0370-2693(01)01295-3. arXiv: hep-ph/0110096.
- [159] C. Armendariz-Picon, T. Damour, and V. Mukhanov. “K-Inflation”. In: *Physics Letters B* 458.2-3 (1999), pages 209–218. doi: 10.1016/S0370-2693(99)00603-6. arXiv: hep-th/9904075.
- [160] Arjun Berera. “Warm Inflation”. In: *Physical Review Letters* 75.18 (1995), pages 3218–3221. doi: 10.1103/PhysRevLett.75.3218. arXiv: astro-ph/9509049.
- [161] Arjun Berera and Li-Zhi Fang. “Thermally Induced Density Perturbations in the Inflation Era”. In: *Physical Review Letters* 74.11 (1995), pages 1912–1915. doi: 10.1103/PhysRevLett.74.1912. arXiv: astro-ph/9501024.
- [162] Arjun Berera. “Interpolating the Stage of Exponential Expansion in the Early Universe: A Possible Alternative with No Reheating”. In: *Physical Review D* 55.6 (1997), pages 3346–3357. doi: 10.1103/PhysRevD.55.3346. arXiv: hep-ph/9612239.
- [163] Arjun Berera. “Warm Inflation in the Adiabatic Regime- a Model, an Existence Proof for Inflationary Dynamics in Quantum Field Theory”. In: *Nuclear Physics B* 585.3 (2000), pages 666–714. doi: 10.1016/S0550-3213(00)00411-9. arXiv: hep-ph/9904409.
- [164] Arjun Berera, Ian G. Moss, and Rudnei O. Ramos. “Warm Inflation and Its Microphysical Basis”. In: *Reports on Progress in Physics* 72.2 (2009), page 026901. doi: 10.1088/0034-4885/72/2/026901. arXiv: 0808.1855 [hep-ph].
- [165] Vahid Kamali, Meysam Motaharfar, and Rudnei O. Ramos. “Recent Developments in Warm Inflation”. In: *Universe* 9.3 (Feb. 2023), page 124. doi: 10.3390/universe9030124. arXiv: 2302.02827 [hep-ph].
- [166] Arjun Berera. “The Warm Inflation Story”. In: *Universe* 9.6 (June 2023), page 272. doi: 10.3390/universe9060272. arXiv: 2305.10879 [hep-ph].

- [167] David H. Lyth and Andrew R. Liddle. *The Primordial Density Perturbation: Cosmology, Inflation and the Origin of Structure*. 1st edition. Cambridge University Press, June 2009. DOI: 10.1017/CB09780511819209.
- [168] Bruce A. Bassett, Shinji Tsujikawa, and David Wands. “Inflation Dynamics and Reheating”. In: *Reviews of Modern Physics* 78.2 (2006), pages 537–589. DOI: 10.1103/RevModPhys.78.537. arXiv: astro-ph/0507632.
- [169] Rouzbeh Allahverdi et al. “Reheating in Inflationary Cosmology: Theory and Applications”. In: *Annual Review of Nuclear and Particle Science* 60.1 (2010), pages 27–51. DOI: 10.1146/annurev.nucl.012809.104511. arXiv: 1001.2600 [hep-th].
- [170] Mustafa A. Amin et al. “Nonperturbative Dynamics Of Reheating After Inflation: A Review”. In: *International Journal of Modern Physics D* 24.01 (Dec. 2014), page 1530003. DOI: 10.1142/S0218271815300037. arXiv: 1410.3808 [hep-ph].
- [171] Michael S. Turner. “Coherent Scalar-Field Oscillations in an Expanding Universe”. In: *Physical Review D* 28.6 (1983), page 1243. DOI: 10.1103/PhysRevD.28.1243.
- [172] L.F. Abbott, Edward Farhi, and Mark B. Wise. “Particle Production in the New Inflationary Cosmology”. In: *Physics Letters B* 117.1-2 (1982), page 29. DOI: 10.1016/0370-2693(82)90867-X.
- [173] A.D. Dolgov and A.D. Linde. “Baryon Asymmetry in the Inflationary Universe”. In: *Physics Letters B* 116.5 (1982), page 329. DOI: 10.1016/0370-2693(82)90292-1.
- [174] Andreas Albrecht et al. “Reheating an Inflationary Universe”. In: *Physical Review Letters* 48.20 (1982), page 1437. DOI: 10.1103/PhysRevLett.48.1437.
- [175] Juergen Baacke, Katrin Heitmann, and Carsten Patzold. “Nonequilibrium Dynamics of Fermions in a Spatially Homogeneous Scalar Background Field”. In: *Physical Review D* 58.12 (1998), page 125013. DOI: 10.1103/PhysRevD.58.125013. arXiv: hep-ph/9806205.
- [176] G. F. Giudice et al. “Production of Massive Fermions at Preheating and Leptogenesis”. In: *Journal of High Energy Physics* 08.08 (1999), page 014. DOI: 10.1088/1126-6708/1999/08/014. arXiv: hep-ph/9905242.
- [177] Patrick B. Greene and Lev Kofman. “Preheating of Fermions”. In: *Physics Letters B* 448.1-2 (1999), pages 6–12. DOI: 10.1016/S0370-2693(99)00020-9. arXiv: hep-ph/9807339.
- [178] Patrick B. Greene and Lev Kofman. “On the Theory of Fermionic Preheating”. In: *Physical Review D* 62.12 (2000), page 123516. DOI: 10.1103/PhysRevD.62.123516. arXiv: hep-ph/0003018.
- [179] M. Peloso and L. Sorbo. “Preheating of Massive Fermions after Inflation: Analytical Results”. In: *Journal of High Energy Physics* 05.05 (2000), page 016. DOI: 10.1088/1126-6708/2000/05/016. arXiv: hep-ph/0003045.
- [180] Juan Garcia-Bellido and Andrei Linde. “Preheating in Hybrid Inflation”. In: *Physical Review D* 57.10 (1998), pages 6075–6088. DOI: 10.1103/PhysRevD.57.6075. arXiv: hep-ph/9711360.
- [181] Gary Felder et al. “Dynamics of Symmetry Breaking and Tachyonic Preheating”. In: *Physical Review Letters* 87.1 (2001), page 011601. DOI: 10.1103/PhysRevLett.87.011601. arXiv: hep-ph/0012142.

- [182] Md Riajul Haque et al. “Primordial Black Hole Reheating”. In: 108.6 (Sept. 2023), page 063523. doi: 10.1103/PhysRevD.108.063523. arXiv: 2305.10518 [hep-ph].
- [183] Viatcheslav F. Mukhanov and Sergei Winitzki. *Introduction to Quantum Effects in Gravity*. Nachdr. Cambridge: Cambridge Univ. Press, 2010.
- [184] N. D. Birrell and P. C. W. Davies. *Quantum Fields in Curved Space*. 1st edition. Cambridge University Press, 1982. doi: 10.1017/CB09780511622632.
- [185] T.S. Bunch and P.C.W. Davies. “Quantum Field Theory in de Sitter Space: Renormalization by Point-Splitting”. In: *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences* 360.1700 (1978), pages 117–134. doi: 10.1098/rspa.1978.0060.
- [186] *NIST Digital Library of Mathematical Functions*. <https://dlmf.nist.gov/>, Release 1.2.5 of 2025-12-15. F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller, B. V. Saunders, H. S. Cohl, and M. A. McClain, eds. URL: <https://dlmf.nist.gov/>.
- [187] David H. Lyth and David Seery. “Classicality of the Primordial Perturbations”. In: *Physics Letters B* 662.4 (2008), pages 309–313. doi: 10.1016/j.physletb.2008.03.010. arXiv: astro-ph/0607647.
- [188] Leonard Parker. “Quantized Fields and Particle Creation in Expanding Universes. I”. In: *Physical Review* 183.5 (1969), pages 1057–1068. doi: 10.1103/PhysRev.183.1057.
- [189] Leonard Parker. “Quantized Fields and Particle Creation in Expanding Universes. II”. In: *Physical Review D* 3.2 (1971), pages 346–356. doi: 10.1103/PhysRevD.3.346.
- [190] Daniel Baumann. “TASI Lectures on Inflation”. In: (2011). doi: 10.1142/9789814327183_0010. arXiv: 0907.5424 [hep-th].
- [191] Ruth Durrer. *The Cosmic Microwave Background*. 1st edition. Cambridge University Press, 2008. doi: 10.1017/CB09780511817205.
- [192] Michele Maggiore. *Gravitational Waves*. Oxford: Oxford university press, 2017.
- [193] Viatcheslav F. Mukhanov. “Gravitational Instability of the Universe Filled with a Scalar Field”. In: *JETP Lett.* 41 (1985), pages 493–496.
- [194] Misao Sasaki. “Large Scale Quantum Fluctuations in the Inflationary Universe”. In: *Prog. Theor. Phys.* 76 (1986), page 1036. doi: 10.1143/PTP.76.1036.
- [195] Wayne Hu and Scott Dodelson. “Cosmic Microwave Background Anisotropies”. In: *Annual Review of Astronomy and Astrophysics* 40.1 (2002), pages 171–216. doi: 10.1146/annurev.astro.40.060401.093926. arXiv: astro-ph/0110414.
- [196] Paolo Cabella and Marc Kamionkowski. “Theory of Cosmic Microwave Background Polarization”. In: (Mar. 2005). doi: 10.48550/arXiv.astro-ph/0403392. arXiv: astro-ph/0403392.
- [197] Planck Collaboration et al. “Planck Intermediate Results. LVII. Joint Planck LFI and HFI Data Processing”. In: *Astronomy & Astrophysics* 643 (Nov. 2020), A42. doi: 10.1051/0004-6361/202038073. arXiv: 2007.04997 [astro-ph.CO].
- [198] M. Tristram et al. “Cosmological Parameters Derived from the Final (PR4) Planck Data Release”. In: *Astron. Astrophys.* 682 (Feb. 2024), A37. doi: 10.1051/0004-6361/202348015. arXiv: 2309.10034 [astro-ph.CO].
- [199] Thibaut Louis et al. “The Atacama Cosmology Telescope: DR6 Power Spectra, Likelihoods and Λ CDM Parameters”. In: (June 2025). doi: 10.48550/arXiv.2503.14452. arXiv: 2503.14452 [astro-ph.CO].

- [200] DESI Collaboration et al. “DESI DR2 Results II: Measurements of Baryon Acoustic Oscillations and Cosmological Constraints”. In: *Phys. Rev. D* 112.8 (Oct. 2025), page 083515. doi: 10.1103/tr6y-kpc6. arXiv: 2503.14738 [astro-ph.CO].
- [201] BICEP/Keck Collaboration et al. “BICEP / Keck XIII: Improved Constraints on Primordial Gravitational Waves Using Planck, WMAP, and BICEP/Keck Observations through the 2018 Observing Season”. In: *Physical Review Letters* 127.15 (Oct. 2021), page 151301. doi: 10.1103/PhysRevLett.127.151301. arXiv: 2110.00483 [astro-ph.CO].
- [202] M. Tristram et al. “Improved Limits on the Tensor-to-Scalar Ratio Using BICEP and Planck”. In: *Physical Review D* 105.8 (Apr. 2022), page 083524. doi: 10.1103/PhysRevD.105.083524. arXiv: 2112.07961 [astro-ph.CO].
- [203] Giacomo Galloni, Sophie Henrot-Versillé, and Matthieu Tristram. “Robust Constraints on Tensor Perturbations from Cosmological Data: A Comparative Analysis from Bayesian and Frequentist Perspectives”. In: *Phys. Rev. D* 110.6 (Sept. 2024), page 063511. doi: 10.1103/PhysRevD.110.063511. arXiv: 2405.04455 [astro-ph.CO].
- [204] Kevork Abazajian et al. “CMB-S4: Forecasting Constraints on Primordial Gravitational Waves”. In: *Astrophys. J.* 926.1 (Feb. 2022), page 54. doi: 10.3847/1538-4357/ac1596. arXiv: 2008.12619 [astro-ph.CO].
- [205] Keck Array et al. “BICEP2 / Keck Array x: Constraints on Primordial Gravitational Waves Using Planck, WMAP, and New BICEP2/Keck Observations through the 2015 Season”. In: *Physical Review Letters* 121.22 (2018), page 221301. doi: 10.1103/PhysRevLett.121.221301. arXiv: 1810.05216 [astro-ph.CO].
- [206] P. A. R. Ade et al. “BICEP2 I: Detection Of B-mode Polarization at Degree Angular Scales”. In: *Phys. Rev. Lett.* 112.24 (June 2014), page 241101. doi: 10.1103/PhysRevLett.112.241101. arXiv: 1403.3985 [astro-ph.CO].
- [207] John Joseph M. Carrasco, Renata Kallosh, and Andrei Linde. “Cosmological Attractors and Initial Conditions for Inflation”. In: *Physical Review D* 92.6 (Sept. 2015), page 063519. doi: 10.1103/PhysRevD.92.063519. arXiv: 1506.00936 [hep-th].
- [208] Mario Galante et al. “The Unity of Cosmological Attractors”. In: *Physical Review Letters* 114.14 (Apr. 2015), page 141302. doi: 10.1103/PhysRevLett.114.141302. arXiv: 1412.3797 [hep-th].
- [209] David H. Lyth. “What Would We Learn by Detecting a Gravitational Wave Signal in the Cosmic Microwave Background Anisotropy?”. In: *Physical Review Letters* 78.10 (1997), pages 1861–1863. doi: 10.1103/PhysRevLett.78.1861. arXiv: hep-ph/9606387.
- [210] M. S. Turner. “Recovering the Inflationary Potential”. In: *Physical Review D* 48.12 (1993), pages 5539–5545. doi: 10.1103/PhysRevD.48.5539. arXiv: astro-ph/9307035.
- [211] V. Acquaviva et al. “Second-Order Cosmological Perturbations from Inflation”. In: *Nuclear Physics B* 667.1-2 (2003), pages 119–148. doi: 10.1016/S0550-3213(03)00550-9. arXiv: astro-ph/0209156.
- [212] Juan Maldacena. “Non-Gaussian Features of Primordial Fluctuations in Single Field Inflationary Models”. In: *Journal of High Energy Physics* 05.05 (2003), page 013. doi: 10.1088/1126-6708/2003/05/013. arXiv: astro-ph/0210603.
- [213] N. Bartolo et al. “Non-Gaussianity from Inflation: Theory and Observations”. In: *Physics Reports* 402.3-4 (2004), pages 103–266. doi: 10.1016/j.physrep.2004.08.022. arXiv: astro-ph/0406398.

- [214] Xingang Chen. “Primordial Non-Gaussianities from Inflation Models”. In: *Advances in Astronomy* 2010.1 (2010), page 638979. doi: 10.1155/2010/638979. arXiv: 1002.1416 [astro-ph.CO].
- [215] Yi Wang. “Inflation, Cosmic Perturbations and Non-Gaussianities”. In: *Communications in Theoretical Physics* 62.1 (2014), pages 109–166. doi: 10.1088/0253-6102/62/1/19. arXiv: 1303.1523 [hep-th].
- [216] Sébastien Renaux-Petel. “Primordial Non-Gaussianities after Planck 2015: An Introductory Review”. In: *Comptes Rendus. Physique* 16.10 (Dec. 2015), pages 969–985. doi: 10.1016/j.crhy.2015.08.003. arXiv: 1508.06740 [astro-ph.CO].
- [217] Ana Achúcarro et al. “Inflation: Theory and Observations”. In: (Sept. 2022). doi: 10.48550/arXiv.2203.08128. arXiv: 2203.08128 [astro-ph.CO].
- [218] Maria Chiara Guzzetti et al. “Gravitational Waves from Inflation”. In: *Riv. Nuovo Cim.* 39.9 (Aug. 2016), pages 399–495. doi: 10.1393/ncr/i2016-10127-1. arXiv: 1605.01615 [astro-ph.CO].
- [219] Chiara Caprini and Daniel G. Figueroa. “Cosmological Backgrounds of Gravitational Waves”. In: *Class. Quant. Grav.* 35.16 (July 2018), page 163001. doi: 10.1088/1361-6382/aac608. arXiv: 1801.04268 [astro-ph.CO].
- [220] Maura McLaughlin. “The North American Nanohertz Observatory for Gravitational Waves”. In: *Classical and Quantum Gravity* 30.22 (Nov. 2013), page 224008. doi: 10.1088/0264-9381/30/22/224008. arXiv: 1310.0758 [astro-ph.IM].
- [221] Michael Kramer and David J Champion. “The European Pulsar Timing Array and the Large European Array for Pulsars”. In: *Classical and Quantum Gravity* 30.22 (Nov. 2013), page 224009. doi: 10.1088/0264-9381/30/22/224009.
- [222] C. Carilli and S. Rawlings. “Science with the Square Kilometer Array: Motivation, Key Science Projects, Standards and Assumptions”. In: *New Astronomy Reviews* 48.11-12 (Dec. 2004), pages 979–984. doi: 10.1016/j.newar.2004.09.001. arXiv: astro-ph/0409274.
- [223] Gregory M Harry and the LIGO Scientific Collaboration. “Advanced LIGO: The next Generation of Gravitational Wave Detectors”. In: *Classical and Quantum Gravity* 27.8 (Apr. 2010), page 084006. doi: 10.1088/0264-9381/27/8/084006.
- [224] The LIGO Scientific Collaboration. “Advanced LIGO”. In: (2014). doi: 10.48550/ARXIV.1411.4547.
- [225] F. Acernese et al. “Advanced Virgo: A 2nd Generation Interferometric Gravitational Wave Detector”. In: (2014). doi: 10.48550/ARXIV.1408.3978.
- [226] Kentaro Somiya. “Detector Configuration of KAGRA - the Japanese Cryogenic Gravitational-Wave Detector”. In: *Classical and Quantum Gravity* 29.12 (2012), page 124007. doi: 10.1088/0264-9381/29/12/124007. arXiv: 1111.7185 [gr-qc].
- [227] Yoichi Aso et al. “Interferometer Design of the KAGRA Gravitational Wave Detector”. In: *Physical Review D* 88.4 (Aug. 2013), page 043007. doi: 10.1103/PhysRevD.88.043007. arXiv: 1306.6747 [gr-qc].
- [228] Pau Amaro-Seoane et al. “Laser Interferometer Space Antenna”. In: (Feb. 2017). doi: 10.48550/arXiv.1702.00786. arXiv: 1702.00786 [astro-ph.IM].

- [229] Vincent Corbin and Neil J. Cornish. “Detecting the Cosmic Gravitational Wave Background with the Big Bang Observer”. In: *Classical and Quantum Gravity* 23.7 (Apr. 2006), pages 2435–2446. doi: 10.1088/0264-9381/23/7/014. arXiv: gr-qc/0512039.
- [230] Naoki Seto, Seiji Kawamura, and Takashi Nakamura. “Possibility of Direct Measurement of the Acceleration of the Universe Using 0.1 Hz Band Laser Interferometer Gravitational Wave Antenna in Space”. In: *Physical Review Letters* 87.22 (Nov. 2001), page 221103. doi: 10.1103/PhysRevLett.87.221103. arXiv: astro-ph/0108011.
- [231] Masaki Satoh and Jiro Soda. “Higher Curvature Corrections to Primordial Fluctuations in Slow-roll Inflation”. In: *Journal of Cosmology and Astroparticle Physics* 09.09 (2008), page 019. doi: 10.1088/1475-7516/2008/09/019. arXiv: 0806.4594 [astro-ph].
- [232] Tomohiro Fujita et al. “Blue-Tilted Primordial Gravitational Waves from Massive Gravity”. In: *Physics Letters B* 789 (Feb. 2019), pages 215–219. doi: 10.1016/j.physletb.2018.12.025. arXiv: 1808.02381 [gr-qc].
- [233] Yun-Song Piao and Yuan-Zhong Zhang. “Phantom Inflation and Primordial Perturbation Spectrum”. In: *Physical Review D* 70.6 (2004), page 063513. doi: 10.1103/PhysRevD.70.063513. arXiv: astro-ph/0401231.
- [234] Tsutomu Kobayashi, Masahide Yamaguchi, and Jun’ichi Yokoyama. “G-Inflation: Inflation Driven by the Galileon Field”. In: *Physical Review Letters* 105.23 (2010), page 231302. doi: 10.1103/PhysRevLett.105.231302. arXiv: 1008.0603 [hep-th].
- [235] Solomon Endlich, Alberto Nicolis, and Junpu Wang. “Solid Inflation”. In: *Journal of Cosmology and Astroparticle Physics* 10.10 (Oct. 2013), page 011. doi: 10.1088/1475-7516/2013/10/011. arXiv: 1210.0569 [hep-th].
- [236] Sachiko Kuroyanagi, Tomo Takahashi, and Shuichiro Yokoyama. “Blue-Tilted Tensor Spectrum and Thermal History of the Universe”. In: *Journal of Cosmology and Astroparticle Physics* 2015.02 (Feb. 2015), pages 003–003. doi: 10.1088/1475-7516/2015/02/003. arXiv: 1407.4785 [astro-ph].
- [237] Daniel G. Figueroa and Erwin H. Tanin. “Ability of LIGO and LISA to Probe the Equation of State of the Early Universe”. In: *Journal of Cosmology and Astroparticle Physics* 2019.08 (Aug. 2019), pages 011–011. doi: 10.1088/1475-7516/2019/08/011. arXiv: 1905.11960 [astro-ph.CO].
- [238] Hannah Duval et al. “Investigating Cosmic Histories with a Stiff Era through Gravitational Waves”. In: (Nov. 2024). doi: 10.48550/arXiv.2405.10201. arXiv: 2405.10201 [gr-qc].
- [239] The LIGO Scientific Collaboration et al. “Upper Limits on the Isotropic Gravitational-Wave Background from Advanced LIGO’s and Advanced Virgo’s Third Observing Run”. In: (2021). doi: 10.48550/ARXIV.2101.12130.
- [240] The LIGO Scientific Collaboration et al. “Upper Limits on the Isotropic Gravitational-Wave Background from the First Part of LIGO, Virgo, and KAGRA’s Fourth Observing Run”. In: (Aug. 2025). doi: 10.48550/arXiv.2508.20721. arXiv: 2508.20721 [gr-qc].
- [241] The LIGO Scientific Collaboration et al. “Cosmological and High Energy Physics Implications from Gravitational-Wave Background Searches in LIGO-Virgo-KAGRA’s O1-O4a Runs”. In: (Nov. 2025). doi: 10.48550/arXiv.2510.26848. arXiv: 2510.26848 [gr-qc].

- [242] Gabriella Agazie et al. “The NANOGrav 15 Yr Data Set: Evidence for a Gravitational-wave Background”. In: *The Astrophysical Journal Letters* 951.1 (July 2023), page L8. doi: 10.3847/2041-8213/acdac6.
- [243] EPTA Collaboration and InPTA Collaboration: et al. “The Second Data Release from the European Pulsar Timing Array: III. Search for Gravitational Wave Signals”. In: *Astronomy & Astrophysics* 678 (Oct. 2023), A50. doi: 10.1051/0004-6361/202346844.
- [244] Heng Xu et al. “Searching for the Nano-Hertz Stochastic Gravitational Wave Background with the Chinese Pulsar Timing Array Data Release I”. In: (2023). doi: 10.48550/ARXIV.2306.16216.
- [245] Andrew Zic et al. “The Parkes Pulsar Timing Array Third Data Release”. In: *Publications of the Astronomical Society of Australia* 40 (2023), e049. doi: 10.1017/pasa.2023.36.
- [246] Matthew T. Miles et al. “The MeerKAT Pulsar Timing Array: The First Search for Gravitational Waves with the MeerKAT Radio Telescope”. In: (2024). doi: 10.48550/ARXIV.2412.01153.
- [247] Kishore N. Ananda, Chris Clarkson, and David Wands. “The Cosmological Gravitational Wave Background from Primordial Density Perturbations”. In: *Physical Review D* 75.12 (2007), page 123518. doi: 10.1103/PhysRevD.75.123518. arXiv: gr-qc/0612013.
- [248] Daniel Baumann et al. “Gravitational Wave Spectrum Induced by Primordial Scalar Perturbations”. In: *Physical Review D* 76.8 (2007), page 084019. doi: 10.1103/PhysRevD.76.084019. arXiv: hep-th/0703290.
- [249] Guillem Domenech. “Scalar Induced Gravitational Waves Review”. In: *Universe* 7.11 (Oct. 2021), page 398. doi: 10.3390/universe7110398. arXiv: 2109.01398 [gr-qc].
- [250] Ya. B. Zel’dovich and I. D. Novikov. “The Hypothesis of Cores Retarded during Expansion and the Hot Cosmological Model”. In: *Sov. Astron.* 10 (1967), page 602.
- [251] Stephen Hawking. “Gravitationally collapsed objects of very low mass”. In: *Mon. Not. Roy. Astron. Soc.* 152 (1971), page 75. doi: 10.1093/mnras/152.1.75.
- [252] B. J. Carr and M. J. Rees. “Can pregalactic objects generate galaxies?” In: *Mon. Not. Roy. Astron. Soc.* 206.4 (1984), pages 801–818. doi: 10.1093/mnras/206.4.801.
- [253] Ogan Özsoy and Gianmassimo Tasinato. “Inflation and Primordial Black Holes”. In: *Universe* 9.5 (Apr. 2023), page 203. doi: 10.3390/universe9050203. arXiv: 2301.03600 [astro-ph.CO].
- [254] Bernard Carr and Florian Kuhnel. “Primordial Black Holes as Dark Matter: Recent Developments”. In: *Annual Review of Nuclear and Particle Science* 70.1 (Aug. 2020), pages 355–394. doi: 10.1146/annurev-nucl-050520-125911. arXiv: 2006.02838 [astro-ph.CO].
- [255] Anne M. Green and Bradley J. Kavanagh. “Primordial Black Holes as a Dark Matter Candidate”. In: *Journal of Physics G: Nuclear and Particle Physics* 48.4 (Feb. 2021), page 043001. doi: 10.1088/1361-6471/abc534. arXiv: 2007.10722 [astro-ph.CO].
- [256] Bernard Carr et al. “Constraints on Primordial Black Holes”. In: *Reports on Progress in Physics* 84.11 (Dec. 2021), page 116902. doi: 10.1088/1361-6633/ac1e31. arXiv: 2002.12778 [astro-ph].
- [257] Misao Sasaki et al. “Primordial Black Holes - Perspectives in Gravitational Wave Astronomy -”. In: *Classical and Quantum Gravity* 35.6 (Feb. 2018), page 063001. doi: 10.1088/1361-6382/aaa7b4. arXiv: 1801.05235 [astro-ph.CO].

- [258] Ryo Saito and Jun'ichi Yokoyama. "Gravitational Wave Background as a Probe of the Primordial Black Hole Abundance". In: *Physical Review Letters* 102.16 (2009), page 161101. doi: 10.1103/PhysRevLett.102.161101. arXiv: 0812.4339 [astro-ph].
- [259] Ryo Saito and Jun'ichi Yokoyama. "Gravitational-Wave Constraints on the Abundance of Primordial Black Holes". In: *Progress of Theoretical Physics* 123.5 (2010), pages 867–886. doi: 10.1143/PTP.126.351. arXiv: 0912.5317 [astro-ph.CO].
- [260] Sidney Coleman. *Aspects of Symmetry: Selected Erice Lectures*. Cambridge, U.K.: Cambridge University Press, 1985. doi: 10.1017/CB09780511565045.
- [261] Steven Weinberg. *The quantum theory of fields. Vol. 2: Modern applications*. Cambridge University Press, Aug. 2013. doi: 10.1017/CB09781139644174.
- [262] Stefan Vandoren and Peter van Nieuwenhuizen. "Lectures on instantons". In: (Feb. 2008). arXiv: 0802.1862 [hep-th].
- [263] Mikhail Shifman. *Advanced topics in quantum field theory.: A lecture course*. Cambridge, UK: Cambridge Univ. Press, Feb. 2012. doi: 10.1017/9781108885911.
- [264] R. Bott. "An Application of Morse theory to the topology of Lie groups". In: *Bull. Soc. Math. Fr.* 84 (1956), pages 251–281.
- [265] A. A. Belavin et al. "Pseudoparticle Solutions of the Yang-Mills Equations". In: *Phys. Lett. B* 59 (1975). Edited by J. C. Taylor, pages 85–87. doi: 10.1016/0370-2693(75)90163-X.
- [266] Curtis G. Callan Jr., R. F. Dashen, and David J. Gross. "The Structure of the Gauge Theory Vacuum". In: *Phys. Lett. B* 63 (1976). Edited by J. C. Taylor, pages 334–340. doi: 10.1016/0370-2693(76)90277-X.
- [267] R. Jackiw and C. Rebbi. "Vacuum Periodicity in a Yang-Mills Quantum Theory". In: *Phys. Rev. Lett.* 37 (1976). Edited by J. C. Taylor, pages 172–175. doi: 10.1103/PhysRevLett.37.172.
- [268] Gerard 't Hooft. "Computation of the Quantum Effects Due to a Four-Dimensional Pseudoparticle". In: *Phys. Rev. D* 14 (1976). Edited by Mikhail A. Shifman. [Erratum: *Phys.Rev.D* 18, 2199 (1978)], pages 3432–3450. doi: 10.1103/PhysRevD.14.3432.
- [269] Curtis G. Callan Jr., Roger F. Dashen, and David J. Gross. "Toward a Theory of the Strong Interactions". In: *Phys. Rev. D* 17 (1978). Edited by Mikhail A. Shifman, page 2717. doi: 10.1103/PhysRevD.17.2717.
- [270] Cumrun Vafa and Edward Witten. "Parity Conservation in QCD". In: *Phys. Rev. Lett.* 53 (1984), page 535. doi: 10.1103/PhysRevLett.53.535.
- [271] C. Vafa and Edward Witten. "Restrictions on Symmetry Breaking in Vector-Like Gauge Theories". In: *Nucl. Phys. B* 234 (1984), pages 173–188. doi: 10.1016/0550-3213(84)90230-X.
- [272] Claude W. Bernard. "Gauge Zero Modes, Instanton Determinants, and QCD Calculations". In: *Phys. Rev. D* 19 (1979). Edited by Mikhail A. Shifman, page 3013. doi: 10.1103/PhysRevD.19.3013.
- [273] David J. Gross, Robert D. Pisarski, and Laurence G. Yaffe. "QCD and Instantons at Finite Temperature". In: *Rev. Mod. Phys.* 53 (1981), page 43. doi: 10.1103/RevModPhys.53.43.
- [274] Michael E. Peskin and Daniel V. Schroeder. *An Introduction to quantum field theory*. Reading, USA: Addison-Wesley, 1995. doi: 10.1201/9780429503559.

- [275] Kazuo Fujikawa. “Path Integral Measure for Gauge Invariant Fermion Theories”. In: *Phys. Rev. Lett.* 42 (1979), pages 1195–1198. doi: 10.1103/PhysRevLett.42.1195.
- [276] Stephen L. Adler. “Axial vector vertex in spinor electrodynamics”. In: *Phys. Rev.* 177 (1969), pages 2426–2438. doi: 10.1103/PhysRev.177.2426.
- [277] J. S. Bell and R. Jackiw. “A PCAC puzzle: $\pi^0 \rightarrow \gamma\gamma$ in the σ model”. In: *Nuovo Cim. A* 60 (1969), pages 47–61. doi: 10.1007/BF02823296.
- [278] S. Navas et al. “Review of particle physics”. In: *Phys. Rev. D* 110.3 (2024), page 030001. doi: 10.1103/PhysRevD.110.030001.
- [279] Varouzhan Baluni. “CP Violating Effects in QCD”. In: *Phys. Rev. D* 19 (1979), pages 2227–2230. doi: 10.1103/PhysRevD.19.2227.
- [280] R. J. Crewther et al. “Chiral Estimate of the Electric Dipole Moment of the Neutron in Quantum Chromodynamics”. In: *Phys. Lett. B* 88 (1979). [Erratum: *Phys.Lett.B* 91, 487 (1980)], page 123. doi: 10.1016/0370-2693(79)90128-X.
- [281] Antonio Pich and Eduardo de Rafael. “Strong CP violation in an effective chiral Lagrangian approach”. In: *Nucl. Phys. B* 367 (1991), pages 313–333. doi: 10.1016/0550-3213(91)90019-T.
- [282] Maxim Pospelov and Adam Ritz. “Theta vacua, QCD sum rules, and the neutron electric dipole moment”. In: *Nucl. Phys. B* 573 (2000), pages 177–200. doi: 10.1016/S0550-3213(99)00817-2. arXiv: hep-ph/9908508.
- [283] M. Abramczyk et al. “Lattice calculation of electric dipole moments and form factors of the nucleon”. In: *Phys. Rev. D* 96.1 (2017), page 014501. doi: 10.1103/PhysRevD.96.014501. arXiv: 1701.07792 [hep-lat].
- [284] Jack Dragos et al. “Confirming the Existence of the strong CP Problem in Lattice QCD with the Gradient Flow”. In: *Phys. Rev. C* 103.1 (2021), page 015202. doi: 10.1103/PhysRevC.103.015202. arXiv: 1902.03254 [hep-lat].
- [285] C. Alexandrou et al. “Neutron electric dipole moment using lattice QCD simulations at the physical point”. In: *Phys. Rev. D* 103.5 (2021), page 054501. doi: 10.1103/PhysRevD.103.054501. arXiv: 2011.01084 [hep-lat].
- [286] Tanmoy Bhattacharya et al. “Contribution of the QCD Θ -term to the nucleon electric dipole moment”. In: *Phys. Rev. D* 103.11 (2021), page 114507. doi: 10.1103/PhysRevD.103.114507. arXiv: 2101.07230 [hep-lat].
- [287] Jian Liang et al. “Nucleon electric dipole moment from the θ term with lattice chiral fermions”. In: *Phys. Rev. D* 108.9 (2023), page 094512. doi: 10.1103/PhysRevD.108.094512. arXiv: 2301.04331 [hep-lat].
- [288] M. A. B. Beg and H. -S. Tsao. “Strong P, T Noninvariances in a Superweak Theory”. In: *Phys. Rev. Lett.* 41 (1978), page 278. doi: 10.1103/PhysRevLett.41.278.
- [289] Howard Georgi. “A Model of Soft CP Violation”. In: *Hadronic J.* 1 (1978), page 155.
- [290] Rabindra N. Mohapatra and G. Senjanovic. “Natural Suppression of Strong p and t Noninvariance”. In: *Phys. Lett. B* 79 (1978), pages 283–286. doi: 10.1016/0370-2693(78)90243-5.
- [291] Ann E. Nelson. “Naturally Weak CP Violation”. In: *Phys. Lett. B* 136 (1984), pages 387–391. doi: 10.1016/0370-2693(84)92025-2.
- [292] Stephen M. Barr. “Solving the Strong CP Problem Without the Peccei-Quinn Symmetry”. In: *Phys. Rev. Lett.* 53 (1984), page 329. doi: 10.1103/PhysRevLett.53.329.

- [293] Luca Vecchi. “Spontaneous CP violation and the strong CP problem”. In: *JHEP* 04 (2017), page 149. doi: 10.1007/JHEP04(2017)149. arXiv: 1412.3805 [hep-ph].
- [294] Michael Dine and Patrick Draper. “Challenges for the Nelson-Barr Mechanism”. In: *JHEP* 08 (2015), page 132. doi: 10.1007/JHEP08(2015)132. arXiv: 1506.05433 [hep-ph].
- [295] S. Yu. Khlebnikov and M. E. Shaposhnikov. “Extra Space-time Dimensions: Towards a Solution to the Strong CP Problem”. In: *Phys. Lett. B* 203 (1988), pages 121–124. doi: 10.1016/0370-2693(88)91582-1.
- [296] Masud Chaichian and Archil. B. Kobakhidze. “Extra dimensions and the strong CP problem”. In: *Phys. Rev. Lett.* 87 (2001), page 171601. doi: 10.1103/PhysRevLett.87.171601. arXiv: hep-ph/0104158.
- [297] S. Khlebnikov and M. Shaposhnikov. “Brane-worlds and theta-vacua”. In: *Phys. Rev. D* 71 (2005), page 104024. doi: 10.1103/PhysRevD.71.104024. arXiv: hep-th/0412306.
- [298] F. L. Bezrukov and Y. Burnier. “Towards a solution of the strong CP problem by compact extra dimensions”. In: *Phys. Rev. D* 80 (2009), page 125004. doi: 10.1103/PhysRevD.80.125004. arXiv: 0811.1163 [hep-th].
- [299] Stuart Samuel. “Does the Yang-Mills theory solve the strong CP problem by itself?” In: *Mod. Phys. Lett. A* 7 (1992), pages 2007–2016. doi: 10.1142/S0217732392001737.
- [300] N. J. Dowrick and N. A. McDougall. “On the Samuel solution to the strong CP problem within QCD”. In: *Nucl. Phys. B* 399 (1993), pages 426–440. doi: 10.1016/0550-3213(93)90503-H.
- [301] Wen-Yuan Ai et al. “Consequences of the order of the limit of infinite spacetime volume and the sum over topological sectors for CP violation in the strong interactions”. In: *Phys. Lett. B* 822 (2021), page 136616. doi: 10.1016/j.physletb.2021.136616. arXiv: 2001.07152 [hep-th].
- [302] Joshua N. Benabou et al. “Clearing up the Strong \$CP\$ Problem”. In: (Oct. 2025). doi: 10.48550/arXiv.2510.18951. arXiv: 2510.18951 [hep-ph].
- [303] Wen-Yuan Ai, Björn Garbrecht, and Carlos Tamarit. “Reply to “Clearing up the Strong \$CP\$ Problem””. In: (Nov. 2025). doi: 10.48550/arXiv.2511.04216. arXiv: 2511.04216 [hep-ph].
- [304] T. W. Donnelly et al. “Do Axions Exist?” In: *Phys. Rev. D* 18 (1978), page 1607. doi: 10.1103/PhysRevD.18.1607.
- [305] P. Di Vecchia and G. Veneziano. “Chiral Dynamics in the Large n Limit”. In: *Nucl. Phys. B* 171 (1980), pages 253–272. doi: 10.1016/0550-3213(80)90370-3.
- [306] Giovanni Grilli di Cortona et al. “The QCD axion, precisely”. In: *JHEP* 01 (2016), page 034. doi: 10.1007/JHEP01(2016)034. arXiv: 1511.02867 [hep-ph].
- [307] P. Sikivie. “Experimental Tests of the Invisible Axion”. In: *Phys. Rev. Lett.* 51 (1983). Edited by M. A. Srednicki. [Erratum: *Phys.Rev.Lett.* 52, 695 (1984)], pages 1415–1417. doi: 10.1103/PhysRevLett.51.1415.
- [308] K. Van Bibber et al. “Proposed experiment to produce and detect light pseudoscalars”. In: *Phys. Rev. Lett.* 59 (1987), pages 759–762. doi: 10.1103/PhysRevLett.59.759.
- [309] Eduard Masso and Ramon Toldra. “New constraints on a light spinless particle coupled to photons”. In: *Phys. Rev. D* 55 (1997), pages 7967–7969. doi: 10.1103/PhysRevD.55.7967. arXiv: hep-ph/9702275.

- [310] Pierluca Carenza et al. “Axion astrophysics”. In: *Phys. Rept.* 1117 (2025), pages 1–102. doi: 10.1016/j.physrep.2025.02.002. arXiv: 2411.02492 [hep-ph].
- [311] David H. Weinberg et al. “Cold dark matter: controversies on small scales”. In: *Proc. Nat. Acad. Sci.* 112 (2015), pages 12249–12255. doi: 10.1073/pnas.1308716112. arXiv: 1306.0913 [astro-ph.CO].
- [312] James S. Bullock and Michael Boylan-Kolchin. “Small-Scale Challenges to the Λ CDM Paradigm”. In: *Ann. Rev. Astron. Astrophys.* 55 (2017), pages 343–387. doi: 10.1146/annurev-astro-091916-055313. arXiv: 1707.04256 [astro-ph.CO].
- [313] Keir K. Rogers and Hiranya V. Peiris. “Strong Bound on Canonical Ultralight Axion Dark Matter from the Lyman-Alpha Forest”. In: *Phys. Rev. Lett.* 126.7 (2021), page 071302. doi: 10.1103/PhysRevLett.126.071302. arXiv: 2007.12705 [astro-ph.CO].
- [314] Keir K. Rogers et al. “Ultra-light axions and the S_8 tension: joint constraints from the cosmic microwave background and galaxy clustering”. In: *JCAP* 06 (2023), page 023. doi: 10.1088/1475-7516/2023/06/023. arXiv: 2301.08361 [astro-ph.CO].
- [315] Andrei Khmelnitsky and Valery Rubakov. “Pulsar Timing Signal from Ultralight Scalar Dark Matter”. In: *JCAP* 02.02 (Feb. 2014), page 019. doi: 10.1088/1475-7516/2014/02/019. arXiv: 1309.5888 [astro-ph.CO].
- [316] Adeela Afzal et al. “The NANOGrav 15 yr Data Set: Search for Signals from New Physics”. In: *Astrophys. J. Lett.* 951.1 (2023). [Erratum: *Astrophys.J.Lett.* 971, L27 (2024), Erratum: *Astrophys.J.* 971, L27 (2024)], page L11. doi: 10.3847/2041-8213/acdc91. arXiv: 2306.16219 [astro-ph.HE].
- [317] Clemente Smarra et al. “Second Data Release from the European Pulsar Timing Array: Challenging the Ultralight Dark Matter Paradigm”. In: *Phys. Rev. Lett.* 131.17 (2023), page 171001. doi: 10.1103/PhysRevLett.131.171001. arXiv: 2306.16228 [astro-ph.HE].
- [318] Clemente Smarra et al. “Constraints on conformal ultralight dark matter couplings from the European Pulsar Timing Array”. In: *Phys. Rev. D* 110.4 (2024), page 043033. doi: 10.1103/PhysRevD.110.043033. arXiv: 2405.01633 [astro-ph.HE].
- [319] Joshua A. Frieman et al. “Cosmology with ultralight pseudo Nambu-Goldstone bosons”. In: *Phys. Rev. Lett.* 75 (1995), pages 2077–2080. doi: 10.1103/PhysRevLett.75.2077. arXiv: astro-ph/9505060.
- [320] Edmund J. Copeland, M. Sami, and Shinji Tsujikawa. “Dynamics of dark energy”. In: *Int. J. Mod. Phys. D* 15 (2006), pages 1753–1936. doi: 10.1142/S021827180600942X. arXiv: hep-th/0603057.
- [321] R. Penrose. “Gravitational collapse: The role of general relativity”. In: *Riv. Nuovo Cim.* 1 (1969), pages 252–276. doi: 10.1023/A:1016578408204.
- [322] Richard Brito, Vitor Cardoso, and Paolo Pani. “Superradiance: New Frontiers in Black Hole Physics”. In: *Lect. Notes Phys.* 906 (2015), pp.1–237. doi: 10.1007/978-3-319-19000-6. arXiv: 1501.06570 [gr-qc].
- [323] Sean M. Carroll, George B. Field, and Roman Jackiw. “Limits on a Lorentz and Parity Violating Modification of Electrodynamics”. In: *Phys. Rev. D* 41 (1990), page 1231. doi: 10.1103/PhysRevD.41.1231.

- [324] Johannes R. Eskilt and Eiichiro Komatsu. “Improved constraints on cosmic birefringence from the WMAP and Planck cosmic microwave background polarization data”. In: *Phys. Rev. D* 106.6 (2022), page 063503. doi: 10.1103/PhysRevD.106.063503. arXiv: 2205.13962 [astro-ph.CO].
- [325] J. R. Eskilt et al. “COSMOGLOBE DR1 results - II. Constraints on isotropic cosmic birefringence from reprocessed WMAP and Planck LFI data”. In: *Astron. Astrophys.* 679 (2023), A144. doi: 10.1051/0004-6361/202346829. arXiv: 2305.02268 [astro-ph.CO].
- [326] P. Diego-Palazuelos and E. Komatsu. “Cosmic Birefringence from the Atacama Cosmology Telescope Data Release 6”. In: (Sept. 2025). arXiv: 2509.13654 [astro-ph.CO].
- [327] Fred Adams et al. “Natural Inflation: Particle Physics Models, Power Law Spectra for Large Scale Structure, and Constraints from COBE”. In: *Phys. Rev. D* 47.2 (1993), pages 426–455. doi: 10.1103/PhysRevD.47.426. arXiv: hep-ph/9207245.
- [328] Jihn E. Kim. “Weak Interaction Singlet and Strong CP Invariance”. In: *Phys. Rev. Lett.* 43 (1979), page 103. doi: 10.1103/PhysRevLett.43.103.
- [329] Mikhail A. Shifman, A. I. Vainshtein, and Valentin I. Zakharov. “Can Confinement Ensure Natural CP Invariance of Strong Interactions?”. In: *Nucl. Phys. B* 166 (1980), pages 493–506. doi: 10.1016/0550-3213(80)90209-6.
- [330] A. R. Zhitnitsky. “On Possible Suppression of the Axion Hadron Interactions. (In Russian)”. In: *Sov. J. Nucl. Phys.* 31 (1980), page 260.
- [331] Michael Dine, Willy Fischler, and Mark Srednicki. “A Simple Solution to the Strong CP Problem with a Harmless Axion”. In: *Phys. Lett. B* 104 (1981), pages 199–202. doi: 10.1016/0370-2693(81)90590-6.
- [332] Peter Svrcek and Edward Witten. “Axions In String Theory”. In: *JHEP* 06 (2006), page 051. doi: 10.1088/1126-6708/2006/06/051. arXiv: hep-th/0605206.
- [333] Asimina Arvanitaki et al. “String Axiverse”. In: *Phys. Rev. D* 81 (2010), page 123530. doi: 10.1103/PhysRevD.81.123530. arXiv: 0905.4720 [hep-th].
- [334] P. Candelas et al. “Vacuum configurations for superstrings”. In: *Nucl. Phys. B* 258 (1985), pages 46–74. doi: 10.1016/0550-3213(85)90602-9.
- [335] Edward Witten. “Some Properties of O(32) Superstrings”. In: *Phys. Lett. B* 149 (1984), pages 351–356. doi: 10.1016/0370-2693(84)90422-2.
- [336] Peter Adshead and Evangelos I. Sfakianakis. “Fermion Production during and after Axion Inflation”. In: *Journal of Cosmology and Astroparticle Physics* 11.11 (Nov. 2015), page 021. doi: 10.1088/1475-7516/2015/11/021. arXiv: 1508.00891 [hep-ph].
- [337] Peter Adshead, Aike Liu, and Kaloian D. Lozanov. “Production and Backreaction of Massive Fermions during Axion Inflation with Non-Abelian Gauge Fields”. In: *Journal of Cosmology and Astroparticle Physics* 09.09 (Sept. 2022), page 043. doi: 10.1088/1475-7516/2022/09/043. arXiv: 2203.09370 [hep-ph].
- [338] Peter Adshead, Emil Martinec, and Mark Wyman. “Perturbations in Chromo-Natural Inflation”. In: *Journal of High Energy Physics* 09.9 (2013), page 087. doi: 10.1007/JHEP09(2013)087. arXiv: 1305.2930 [hep-th].
- [339] Neil Barnaby, Enrico Pajer, and Marco Peloso. “Gauge Field Production in Axion Inflation: Consequences for Monodromy, Non-Gaussianity in the CMB, and Gravitational Waves at Interferometers”. In: *Phys. Rev. D* 85.2 (2012), page 023525. doi: 10.1103/PhysRevD.85.023525. arXiv: 1110.3327 [astro-ph.CO].

- [340] Kai-Ge Zhang et al. *Suppression of Scalar Perturbations Due to a Heavy Axion*. Oct. 2025. doi: 10.48550/arXiv.2510.02828. arXiv: 2510.02828 [hep-ph].
- [341] William H. Kinney and K. T. Mahanthappa. “Natural Inflation From Fermion Loops”. In: *Physical Review D* 52.10 (1995), pages 5529–5537. doi: 10.1103/PhysRevD.52.5529. arXiv: hep-ph/9503331.
- [342] William H. Kinney and K. T. Mahanthappa. “Inflation at Low Scales: General Analysis and a Detailed Model”. In: *Physical Review D* 53.10 (1996), pages 5455–5467. doi: 10.1103/PhysRevD.53.5455. arXiv: hep-ph/9512241.
- [343] R. Holman et al. “Solutions to the Strong CP Problem in a World with Gravity”. In: *Physics Letters B* 282.1-2 (1992), pages 132–136. doi: 10.1016/0370-2693(92)90491-L. arXiv: hep-ph/9203206.
- [344] Marc Kamionkowski and John March-Russell. “Planck-Scale Physics and the Peccei-Quinn Mechanism”. In: *Physics Letters B* 282.1-2 (1992), pages 137–141. doi: 10.1016/0370-2693(92)90492-M. arXiv: hep-th/9202003.
- [345] Renata Kallosh et al. “Gravity and Global Symmetries”. In: *Physical Review D* 52.2 (1995), pages 912–935. doi: 10.1103/PhysRevD.52.912. arXiv: hep-th/9502069.
- [346] Tom Banks et al. “On the Possibility of Large Axion Decay Constants”. In: *Journal of Cosmology and Astroparticle Physics* 06.06 (2003), page 001. doi: 10.1088/1475-7516/2003/06/001. arXiv: hep-th/0303252.
- [347] Y. Akrami et al. “Planck 2018 results. X. Constraints on inflation”. In: *Astron. Astrophys.* 641 (2020), A10. doi: 10.1051/0004-6361/201833887. arXiv: 1807.06211 [astro-ph.CO].
- [348] Xi Dong et al. “Simple Exercises to Flatten Your Potential”. In: *Physical Review D* 84.2 (2011), page 026011. doi: 10.1103/PhysRevD.84.026011. arXiv: 1011.4521 [hep-th].
- [349] Guido D’Amico, Nemanja Kaloper, and Albion Lawrence. “Monodromy Inflation at Strong Coupling: $4\pi f$ in the Sky”. In: *Physical Review Letters* 121.9 (Aug. 2018), page 091301. doi: 10.1103/PhysRevLett.121.091301. arXiv: 1709.07014 [hep-th].
- [350] Raphael Flauger et al. “Oscillations in the CMB from Axion Monodromy Inflation”. In: *Journal of Cosmology and Astroparticle Physics* 06.06 (2010), page 009. doi: 10.1088/1475-7516/2010/06/009. arXiv: 0907.2916 [hep-th].
- [351] Angelo Caravano, Keisuke Inomata, and Sébastien Renaux-Petel. “Inflationary Butterfly Effect: Non-perturbative Dynamics From Small-Scale Features”. In: *Physical Review Letters* 133.15 (Oct. 2024), page 151001. doi: 10.1103/PhysRevLett.133.151001. arXiv: 2403.12811 [astro-ph.CO].
- [352] Marco Peloso and Caner Unal. “Trajectories with Suppressed Tensor-to-Scalar Ratio in Aligned Natural Inflation”. In: *Journal of Cosmology and Astroparticle Physics* 06.06 (June 2015), page 040. doi: 10.1088/1475-7516/2015/06/040. arXiv: 1504.02784 [astro-ph.CO].
- [353] Federico Greco and Marco Peloso. “Analytic Results in Aligned Axion Inflation”. In: *JCAP* 01 (Jan. 2025), page 074. doi: 10.1088/1475-7516/2025/01/074. arXiv: 2409.01126 [astro-ph.CO].
- [354] Richard Easter and Liam McAllister. “Random Matrices and the Spectrum of N-flation”. In: *Journal of Cosmology and Astroparticle Physics* 05.05 (2006), page 018. doi: 10.1088/1475-7516/2006/05/018. arXiv: hep-th/0512102.

- [355] Soo A. Kim, Andrew R. Liddle, and David Seery. “Non-Gaussianity in Axion Nflation Models”. In: *Physical Review Letters* 105.18 (2010), page 181302. doi: 10.1103/PhysRevLett.105.181302. arXiv: 1005.4410 [astro-ph.CO].
- [356] Soo A. Kim, Andrew R. Liddle, and David Seery. “Non-Gaussianity in Axion Nflation Models: Detailed Predictions and Mass Spectra”. In: *Physical Review D* 85.2 (2012), page 023532. doi: 10.1103/PhysRevD.85.023532. arXiv: 1108.2944 [astro-ph.CO].
- [357] Thomas W. Grimm. “Axion Inflation in Type II String Theory”. In: *Physical Review D* 77.12 (2008), page 126007. doi: 10.1103/PhysRevD.77.126007. arXiv: 0710.3883 [hep-th].
- [358] Renata Kallosh, Navin Sivanandam, and Masoud Soroush. “Axion Inflation and Gravity Waves in String Theory”. In: *Physical Review D* 77.4 (2008), page 043501. doi: 10.1103/PhysRevD.77.043501. arXiv: 0710.3429 [hep-th].
- [359] Peter Adshead and Kaloian D. Lozanov. “Axion Anomalies”. In: *Journal of High Energy Physics* 08.8 (Aug. 2022), page 077. doi: 10.1007/JHEP08(2022)077. arXiv: 2112.07645 [hep-th].
- [360] Emanuela Dimastrogiovanni and Marco Peloso. “Stability Analysis of Chromo-Natural Inflation and Possible Evasion of Lyth’s Bound”. In: *Physical Review D* 87.10 (May 2013), page 103501. doi: 10.1103/PhysRevD.87.103501. arXiv: 1212.5184 [astro-ph.CO].
- [361] Peter Adshead, Emil Martinec, and Mark Wyman. “Gauge Fields and Inflation: Chiral Gravitational Waves, Fluctuations and the Lyth Bound”. In: *Physical Review D* 88.2 (July 2013), page 021302. doi: 10.1103/PhysRevD.88.021302. arXiv: 1301.2598 [hep-th].
- [362] Valerie Domcke et al. “Emerging Chromo-Natural Inflation”. In: *Journal of Cosmology and Astroparticle Physics* 04.04 (Apr. 2019), page 034. doi: 10.1088/1475-7516/2019/04/034. arXiv: 1807.03358 [hep-ph].
- [363] A. Maleknejad and M. M. Sheikh-Jabbari. “Non-Abelian Gauge Field Inflation”. In: *Physical Review D* 84.4 (2011), page 043515. doi: 10.1103/PhysRevD.84.043515. arXiv: 1102.1932 [hep-ph].
- [364] A. Maleknejad and M. M. Sheikh-Jabbari. “Gauge-Flation: Inflation From Non-Abelian Gauge Fields”. In: 723 (June 2013), pages 224–228. doi: 10.1016/j.physletb.2013.05.001. arXiv: 1102.1513 [hep-ph].
- [365] Vahid Kamali. “Warm (Pseudo)Scalar Inflation”. In: *Physical Review D* 100.4 (Aug. 2019), page 043520. doi: 10.1103/PhysRevD.100.043520. arXiv: 1901.01897 [gr-qc].
- [366] William DeRocco, Peter W. Graham, and Saarik Kalia. “Warming up Cold Inflation”. In: *Journal of Cosmology and Astroparticle Physics* 2021.11 (Nov. 2021), page 011. doi: 10.1088/1475-7516/2021/11/011. arXiv: 2107.07517 [hep-ph].
- [367] Kim V. Berghaus, Peter W. Graham, and David E. Kaplan. “Minimal Warm Inflation”. In: *Journal of Cosmology and Astroparticle Physics* 2020.03 (Mar. 2020), pages 034–034. doi: 10.1088/1475-7516/2020/03/034. arXiv: 1910.07525 [hep-ph].
- [368] Kim V. Berghaus, Matthew Forslund, and Mark Vincent Guevarra. “Warm Inflation with a Heavy QCD Axion”. In: *Journal of Cosmology and Astroparticle Physics* 10.10 (Oct. 2024), page 103. doi: 10.1088/1475-7516/2024/10/103. arXiv: 2402.13535 [hep-ph].

- [369] Sukannya Bhattacharya et al. “On the Prospects of Thermalization of Axion-SU(2) Inflation”. In: *Journal of Cosmology and Astroparticle Physics* 10.10 (Oct. 2025), page 080. DOI: 10.1088/1475-7516/2025/10/080. arXiv: 2506.11853 [astro-ph.CO].
- [370] Ann Mukuno and Jiro Soda. “Chromonatural Warm Inflation”. In: *Physical Review D* 109.12 (June 2024), page 123504. DOI: 10.1103/PhysRevD.109.123504. arXiv: 2402.08849 [hep-th].
- [371] Valerie Domcke and Stefan Sandner. “The Different Regimes of Axion Gauge Field Inflation”. In: *Journal of Cosmology and Astroparticle Physics* 2019.09 (Sept. 2019), pages 038–038. DOI: 10.1088/1475-7516/2019/09/038. arXiv: 1905.11372 [astro-ph.CO].
- [372] Valerie Domcke, Yohei Ema, and Stefan Sandner. “Perturbatively Including Inhomogeneities in Axion Inflation”. In: *JCAP* 03.03 (Mar. 2024), page 019. DOI: 10.1088/1475-7516/2024/03/019. arXiv: 2310.09186 [astro-ph.CO].
- [373] M. Ballardini et al. “Energy-Momentum Tensor and Helicity for Gauge Fields Coupled to a Pseudo-Scalar Inflaton”. In: *Phys. Rev. D* 100.12 (Dec. 2019), page 123542. DOI: 10.1103/PhysRevD.100.123542. arXiv: 1910.13448 [gr-qc].
- [374] Chiara Animalì, Pietro Conzini, and Giovanni Marozzi. “On Adiabatic Renormalization with a Physically Motivated Infrared Cut-Off”. In: *JCAP* 05.05 (May 2022), page 026. DOI: 10.1088/1475-7516/2022/05/026. arXiv: 2201.05602 [gr-qc].
- [375] Mar Bastero-Gil et al. “Dark Photon Dark Matter from a Rolling Inflaton”. In: *JCAP* 02.02 (Feb. 2022), page 015. DOI: 10.1088/1475-7516/2022/02/015. arXiv: 2103.12145 [hep-ph].
- [376] Angelo Caravano et al. “Lattice Simulations of Abelian Gauge Fields Coupled to Axions during Inflation”. In: *Phys. Rev. D* 105.12 (June 2022), page 123530. DOI: 10.1103/PhysRevD.105.123530. arXiv: 2110.10695 [astro-ph.CO].
- [377] Angelo Caravano et al. “Lattice Simulations of Axion-U(1) Inflation”. In: *Phys. Rev. D* 108.4 (Aug. 2023), page 043504. DOI: 10.1103/PhysRevD.108.043504. arXiv: 2204.12874 [astro-ph.CO].
- [378] Daniel G. Figueroa et al. “The Strong Backreaction Regime in Axion Inflation”. In: *Phys. Rev. Lett.* 131.15 (Oct. 2023), page 151003. DOI: 10.1103/PhysRevLett.131.151003. arXiv: 2303.17436 [astro-ph.CO].
- [379] Daniel G. Figueroa et al. “The Non-Linear Dynamics of Axion Inflation: A Detailed Lattice Study”. In: *Phys. Rev. D* 111.6 (Mar. 2025), page 063545. DOI: 10.1103/PhysRevD.111.063545. arXiv: 2411.16368 [astro-ph.CO].
- [380] Ramkishor Sharma et al. “Lattice Simulations of Axion-U(1) Inflation: Gravitational Waves, Magnetic Fields, and Scalar Statistics”. In: *JCAP* 05.05 (May 2025), page 079. DOI: 10.1088/1475-7516/2025/05/079. arXiv: 2411.04854 [astro-ph.CO].
- [381] Drew Jamieson, Angelo Caravano, and Eiichiro Komatsu. “Primordial Power Spectrum and Bispectrum from Lattice Simulations of Axion-U(1) Inflation”. In: (July 2025). DOI: 10.48550/arXiv.2507.22285. arXiv: 2507.22285 [astro-ph.CO].
- [382] Sofia P. Corbà and Lorenzo Sorbo. “Correlated Scalar Perturbations and Gravitational Waves from Axion Inflation”. In: *JCAP* 10 (Oct. 2024), page 024. DOI: 10.1088/1475-7516/2024/10/024. arXiv: 2403.03338 [astro-ph.CO].
- [383] Ricardo Z. Ferreira and Martin S. Sloth. “Universal Constraints on Axions from Inflation”. In: *JHEP* 12.12 (Dec. 2014), page 139. DOI: 10.1007/JHEP12(2014)139. arXiv: 1409.5799 [hep-ph].

- [384] Jun'ya Kume, Marco Peloso, and Nicola Bartolo. “Revisiting the Chern-Simons Interaction during Inflation with a Non-Canonical Pseudo-Scalar”. In: 08 (Aug. 2025), page 044. doi: 10.1088/1475-7516/2025/08/044. arXiv: 2501.02890 [astro-ph.CO].
- [385] Peter Adshead et al. “Constraining Axion Inflation with Gravitational Waves across 29 Decades in Frequency”. In: *Phys. Rev. Lett.* 124.17 (Apr. 2020), page 17. doi: 10.1103/PhysRevLett.124.171301. arXiv: 1909.12843 [astro-ph.CO].
- [386] Peter Adshead et al. “Constraining Axion Inflation with Gravitational Waves from Preheating”. In: *Phys. Rev. D* 101.8 (Apr. 2020), page 8. doi: 10.1103/PhysRevD.101.083534. arXiv: 1909.12842 [astro-ph.CO].
- [387] Peter Adshead et al. “Magnetogenesis from Axion Inflation”. In: *JCAP* 10.10 (Oct. 2016), page 039. doi: 10.1088/1475-7516/2016/10/039. arXiv: 1606.08474 [astro-ph.CO].
- [388] Ricardo Z. Ferreira and Alessio Notari. “Thermalized Axion Inflation”. In: *JCAP* 09.09 (Sept. 2017), page 007. doi: 10.1088/1475-7516/2017/09/007. arXiv: 1706.00373 [astro-ph.CO].
- [389] Ricardo Z. Ferreira and Alessio Notari. “Thermalized Axion Inflation: Natural and Monomial Inflation with Small r ”. In: *Phys. Rev. D* 97.6 (Mar. 2018), page 063528. doi: 10.1103/PhysRevD.97.063528. arXiv: 1711.07483 [astro-ph.CO].
- [390] R. Durrer, O. Sobol, and S. Vilchinskii. “Backreaction from Gauge Fields Produced during Inflation”. In: *Phys. Rev. D* 108.4 (Aug. 2023), page 043540. doi: 10.1103/PhysRevD.108.043540. arXiv: 2303.04583 [gr-qc].
- [391] A. V. Lysenko. “Gradient Expansion Formalism for a Generic Model of Inflationary Magnetogenesis”. In: *Eur. Phys. J. C* 85.5 (May 2025), page 525. doi: 10.1140/epjc/s10052-025-14269-9.
- [392] A. V. Lysenko, O. O. Sobol, and S. I. Vilchinskii. “Vector Dark Matter Production during Inflation in the Gradient-Expansion Formalism”. In: (Oct. 2025). doi: 10.48550/arXiv.2509.24963. arXiv: 2509.24963 [hep-ph].
- [393] Bharat Ratra. “Cosmological ‘seed’ Magnetic Field from Inflation”. In: *Astrophys. J. Lett.* 391 (1992), pages L1–L4. doi: 10.1086/186384.
- [394] Richard von Eckardstein. “GEFF: The Gradient Expansion Formalism Factory - A Tool for Inflationary Gauge-Field Production”. In: (Oct. 2025). doi: 10.48550/arXiv.2510.12644. arXiv: 2510.12644 [astro-ph.CO].
- [395] Daniel G. Figueroa et al. “The Art of Simulating the Early Universe – Part I”. In: *Journal of Cosmology and Astroparticle Physics* 04.04 (2021), page 035. doi: 10.1088/1475-7516/2021/04/035. arXiv: 2006.15122 [astro-ph.CO].
- [396] Jorge Baeza-Ballesteros et al. “The Art of Simulating the Early Universe. Part II. Non-canonical Cases & Gravitational Waves”. In: (Jan. 2026). doi: 10.48550/arXiv.2512.15627. arXiv: 2512.15627 [astro-ph.CO].
- [397] Takahiro Hayashinaka, Tomohiro Fujita, and Jun'ichi Yokoyama. “Fermionic Schwinger Effect and Induced Current in de Sitter Space”. In: *JCAP* 07.07 (July 2016), page 010. doi: 10.1088/1475-7516/2016/07/010. arXiv: 1603.04165 [hep-th].
- [398] Clément Stahl and She-Sheng Xue. “Schwinger Effect and Backreaction in de Sitter Space-time”. In: *Phys. Lett. B* 760 (Sept. 2016), pages 288–292. doi: 10.1016/j.physletb.2016.07.011. arXiv: 1603.07166 [hep-th].

- [399] Ehsan Bavarsad, Clément Stahl, and She-Sheng Xue. “Scalar Current of Created Pairs by Schwinger Mechanism in de Sitter Spacetime”. In: *Phys. Rev. D* 94.10 (Nov. 2016), page 104011. doi: 10.1103/PhysRevD.94.104011. arXiv: 1602.06556 [hep-th].
- [400] Takahiro Hayashinaka and Jun’ichi Yokoyama. “Point Splitting Renormalization of Schwinger Induced Current in de Sitter Spacetime”. In: *JCAP* 07.07 (July 2016), page 012. doi: 10.1088/1475-7516/2016/07/012. arXiv: 1603.06172 [hep-th].
- [401] Ramkishor Sharma and Suprit Singh. “Multifaceted Schwinger Effect in de Sitter Space”. In: *Phys. Rev. D* 96.2 (July 2017), page 025012. doi: 10.1103/PhysRevD.96.025012. arXiv: 1704.05076 [gr-qc].
- [402] Takahiro Hayashinaka and She-Sheng Xue. “Physical Renormalization Condition for de Sitter QED”. In: *Phys. Rev. D* 97.10 (May 2018), page 105010. doi: 10.1103/PhysRevD.97.105010. arXiv: 1802.03686 [gr-qc].
- [403] Karthik Rajeev, Sumanta Chakraborty, and T. Padmanabhan. “Generalized Schwinger Effect and Particle Production in an Expanding Universe”. In: *Phys. Rev. D* 100.4 (Aug. 2019), page 045019. doi: 10.1103/PhysRevD.100.045019. arXiv: 1904.03207 [gr-qc].
- [404] Thomas D. Cohen and David A. McGady. “The Schwinger Mechanism Revisited”. In: *Physical Review D* 78.3 (2008), page 036008. doi: 10.1103/PhysRevD.78.036008. arXiv: 0807.1117 [hep-ph].
- [405] Hagen Kleinert, Remo Ruffini, and She-Sheng Xue. “Electron-Positron Pair Production in Space- or Time-Dependent Electric Fields”. In: *Physical Review D* 78.2 (2008), page 025011. doi: 10.1103/PhysRevD.78.025011. arXiv: 0807.0909 [hep-th].
- [406] A. I. Nikishov. “Pair Production by a Constant External Field”. In: *Zh. Eksp. Teor. Fiz.* 57 (1969), pages 1210–1216.
- [407] A. I. Nikishov. “Barrier Scattering in Field Theory Removal of Klein Paradox”. In: *Nucl. Phys. B* 21 (1970), pages 346–358. doi: 10.1016/0550-3213(70)90527-4.
- [408] Sang Pyo Kim and Don N. Page. “Schwinger Pair Production in Electric and Magnetic Fields”. In: *Physical Review D* 73.6 (2006), page 065020. doi: 10.1103/PhysRevD.73.065020. arXiv: hep-th/0301132.
- [409] Julian Schwinger. “The Theory of Quantized Fields. V”. In: *Physical Review* 93.3 (1954), pages 615–628. doi: 10.1103/PhysRev.93.615.
- [410] E. Brezin and C. Itzykson. “Pair Production in Vacuum by an Alternating Field”. In: *Physical Review D* 2.7 (1970), pages 1191–1199. doi: 10.1103/PhysRevD.2.1191.
- [411] Sang Pyo Kim and Don N. Page. “Improved Approximations for Fermion Pair Production in Inhomogeneous Electric Fields”. In: *Physical Review D* 75.4 (2007), page 045013. doi: 10.1103/PhysRevD.75.045013. arXiv: hep-th/0701047.
- [412] Oliver Gould and Arttu Rajantie. “Thermal Schwinger Pair Production at Arbitrary Coupling”. In: *Physical Review D* 96.7 (Oct. 2017), page 076002. doi: 10.1103/PhysRevD.96.076002. arXiv: 1704.04801 [hep-th].
- [413] Lucas Vicente García-Consuegra and Azadeh Maleknejad. “The Stochastic Schwinger Effect”. In: (Oct. 2025). doi: 10.48550/arXiv.2510.14468. arXiv: 2510.14468 [hep-th].
- [414] Naoto Tanji. “Dynamical View of Pair Creation in Uniform Electric and Magnetic Fields”. In: *Annals of Physics* 324.8 (2009), pages 1691–1736. doi: 10.1016/j.aop.2009.03.012. arXiv: 0810.4429 [hep-ph].

- [415] Harmen J. Warringa. “Dynamics of the Chiral Magnetic Effect in a Weak Magnetic Field”. In: *Physical Review D* 86.8 (2012), page 085029. doi: 10.1103/PhysRevD.86.085029. arXiv: 1205.5679 [hep-th].
- [416] Mariona Banyeres, Guillem Domènech, and Jaume Garriga. “Vacuum Birefringence and the Schwinger Effect in (3+1) de Sitter”. In: *JCAP* 10.10 (Oct. 2018), page 023. doi: 10.1088/1475-7516/2018/10/023. arXiv: 1809.08977 [hep-th].
- [417] Mar Bastero-Gil et al. “Schwinger Current in de Sitter Space”. In: (Mar. 2025). doi: 10.48550/arXiv.2503.01981. arXiv: 2503.01981 [hep-ph].
- [418] Mar Bastero-Gil et al. “Classical Constant Electric Fields and the Schwinger Effect in de Sitter”. In: (Aug. 2025). doi: 10.48550/arXiv.2508.14973. arXiv: 2508.14973 [hep-ph].
- [419] Ramkishor Sharma et al. “Challenges in Inflationary Magnetogenesis: Constraints from Strong Coupling, Backreaction and the Schwinger Effect”. In: *Phys. Rev. D* 96.8 (Oct. 2017), page 083511. doi: 10.1103/PhysRevD.96.083511. arXiv: 1708.08119 [astro-ph.CO].
- [420] O. O. Sobol et al. “Influence of Backreaction of Electric Fields and Schwinger Effect on Inflationary Magnetogenesis”. In: *Phys. Rev. D* 98.6 (Sept. 2018), page 063534. doi: 10.1103/PhysRevD.98.063534. arXiv: 1807.09851 [hep-ph].
- [421] E. V. Gorbar et al. “Hydrodynamical Approach to Chirality Production during Axion Inflation”. In: *Phys. Rev. D* 109.2 (Jan. 2024), page 023536. doi: 10.1103/PhysRevD.109.023536. arXiv: 2311.07429 [hep-ph].
- [422] Michele Maggiore. *A Modern Introduction to Classical Electrodynamics*. Oxford Master Series in Statistical, Computational, and Theoretical Physics. Oxford, United Kingdom: Oxford University Press, 2023.
- [423] M. Srednicki. *Quantum Field Theory*. Cambridge University Press, Jan. 2007. doi: 10.1017/CB09780511813917.
- [424] Hiroyuki Kitamoto. “Schwinger Effect in Inflaton-Driven Electric Field”. In: *Phys. Rev. D* 98.10 (Nov. 2018), page 103512. doi: 10.1103/PhysRevD.98.103512. arXiv: 1807.03753 [hep-th].
- [425] E. V. Gorbar et al. “Kinetic Approach to the Schwinger Effect during Inflation”. In: *Phys. Rev. D* 100.12 (Dec. 2019), page 123502. doi: 10.1103/PhysRevD.100.123502. arXiv: 1909.10332 [gr-qc].
- [426] O. O. Sobol et al. “Schwinger Production of Scalar Particles during and after Inflation from the First Principles”. In: *Phys. Rev. D* 102.2 (July 2020), page 023506. doi: 10.1103/PhysRevD.102.023506. arXiv: 2004.12664 [gr-qc].
- [427] Anastasia V. Lysenko and Oleksandr O. Sobol. “Quantum Kinetic Approach to the Schwinger Production of Scalar Particles in an Expanding Universe”. In: *Gen. Rel. Grav.* 56.4 (Mar. 2024), page 39. doi: 10.1007/s10714-024-03226-8. arXiv: 2311.15981 [hep-ph].
- [428] Arko Bhaumik, Theodoros Papanikolaou, and Anish Ghoshal. “Vector Induced Gravitational Waves Sourced by Primordial Magnetic Fields”. In: (July 2025). doi: 10.48550/arXiv.2504.10477. arXiv: 2504.10477 [astro-ph].
- [429] Ruth Durrer. “Gauge-Invariant Cosmological Perturbation Theory with Seeds”. In: *Physical Review D* 42.8 (1990), pages 2533–2540. doi: 10.1103/PhysRevD.42.2533.

- [430] Ruth Durrer. “Gauge Invariant Cosmological Perturbation Theory”. In: 15 (1994), pages 209–339. doi: 10.48550/arXiv.astro-ph/9311041. arXiv: astro-ph/9311041.
- [431] Shuichiro Yokoyama and Jiro Soda. “Primordial Statistical Anisotropy Generated at the End of Inflation”. In: *JCAP* 08.08 (2008), page 005. doi: 10.1088/1475-7516/2008/08/005. arXiv: 0805.4265 [astro-ph].
- [432] Emanuela Dimastrogiovanni et al. “Non-Gaussianity and Statistical Anisotropy from Vector Field Populated Inflationary Models”. In: *Adv. Astron.* 2010.1 (2010), page 752670. doi: 10.1155/2010/752670. arXiv: 1001.4049 [astro-ph.CO].
- [433] Camille Bonvin, Chiara Caprini, and Ruth Durrer. “Magnetic Fields from Inflation: The Transition to the Radiation Era”. In: *Physical Review D* 86.2 (2012), page 023519. doi: 10.1103/PhysRevD.86.023519. arXiv: 1112.3901 [astro-ph.CO].
- [434] Takahiro T. Nakamura and Ewan D. Stewart. “The Spectrum of Cosmological Perturbations Produced by a Multi-Component Inflaton to Second Order in the Slow-Roll Approximation”. In: *Physics Letters B* 381.4 (1996), pages 413–419. doi: 10.1016/0370-2693(96)00594-1. arXiv: astro-ph/9604103.
- [435] Jinn-Ouk Gong and Takahiro Tanaka. “A Covariant Approach to General Field Space Metric in Multi-Field Inflation”. In: *Journal of Cosmology and Astroparticle Physics* 2011.03 (Mar. 2011), pages 015–015. doi: 10.1088/1475-7516/2011/03/015. arXiv: 1101.4809 [astro-ph.CO].
- [436] Guillermo Ballesteros, Alejandro Pérez Rodríguez, and Mathias Pierre. “Monomial Warm Inflation Revisited”. In: 03 (Mar. 2024), page 003. doi: 10.1088/1475-7516/2024/03/003. arXiv: 2304.05978 [astro-ph.CO].
- [437] A. Gangui et al. “The Three-Point Correlation Function of the Cosmic Microwave Background in Inflationary Models”. In: *The Astrophysical Journal* 430 (Aug. 1994), page 447. doi: 10.1086/174421. arXiv: astro-ph/9312033.
- [438] Daniel Babich, Paolo Creminelli, and Matias Zaldarriaga. “The Shape of Non-Gaussianities”. In: *Journal of Cosmology and Astroparticle Physics* 08.08 (2004), page 009. doi: 10.1088/1475-7516/2004/08/009. arXiv: astro-ph/0405356.
- [439] J. R. Fergusson and E. P. S. Shellard. “The Shape of Primordial Non-Gaussianity and the CMB Bispectrum”. In: *Physical Review D* 80.4 (2009), page 043510. doi: 10.1103/PhysRevD.80.043510. arXiv: 0812.3413 [astro-ph].
- [440] Eiichiro Komatsu and David N. Spergel. “Acoustic Signatures in the Primary Microwave Background Bispectrum”. In: *Physical Review D* 63.6 (2001), page 063002. doi: 10.1103/PhysRevD.63.063002. arXiv: astro-ph/0005036.
- [441] Paolo Creminelli et al. “Limits on Non-Gaussianities from WMAP Data”. In: *Journal of Cosmology and Astroparticle Physics* 05.05 (2006), page 004. doi: 10.1088/1475-7516/2006/05/004. arXiv: astro-ph/0509029.
- [442] Leonardo Senatore, Kendrick M. Smith, and Matias Zaldarriaga. “Non-Gaussianities in Single Field Inflation and Their Optimal Limits from the WMAP 5-Year Data”. In: *Journal of Cosmology and Astroparticle Physics* 01.01 (2010), page 028. doi: 10.1088/1475-7516/2010/01/028. arXiv: 0905.3746 [astro-ph.CO].
- [443] David Seery. “Magnetogenesis and the Primordial Non-Gaussianity”. In: *Journal of Cosmology and Astroparticle Physics* 08.08 (2009), page 018. doi: 10.1088/1475-7516/2009/08/018. arXiv: 0810.1617 [astro-ph].

- [444] Silvia Pla and Ben A. Stefanek. “Renormalization of the Primordial Inflationary Power Spectra”. In: 856 (Aug. 2024), page 138926. doi: 10.1016/j.physletb.2024.138926. arXiv: 2402.14910 [gr-qc].
- [445] Davide Campanella Galanti et al. “Gauge Invariant Quantum Backreaction in U(1) Axion Inflation”. In: *Phys. Rev. D* 110.12 (Dec. 2024), page 123510. doi: 10.1103/PhysRevD.110.123510. arXiv: 2406.19960 [gr-qc].
- [446] Wei Xue, Xian Gao, and Robert Brandenberger. “IR Divergences in Inflation and Entropy Perturbations”. In: *Journal of Cosmology and Astroparticle Physics* 06.06 (2012), page 035. doi: 10.1088/1475-7516/2012/06/035. arXiv: 1201.0768 [hep-th].
- [447] Jason Kristiano and Jun’ichi Yokoyama. “Constraining Primordial Black Hole Formation from Single-Field Inflation”. In: *Physical Review Letters* 132.22 (May 2024), page 221003. doi: 10.1103/PhysRevLett.132.221003. arXiv: 2211.03395 [hep-th].
- [448] A. Riotto. “The Primordial Black Hole Formation from Single-Field Inflation Is Still Not Ruled Out”. In: (Mar. 2023). doi: 10.48550/arXiv.2303.01727. arXiv: 2303.01727 [astro-ph.CO].
- [449] Keisuke Inomata. “Conservation of Superhorizon Curvature Perturbations at One Loop: Backreaction in the in-in Formalism and Renormalization”. In: *Phys. Rev. D* 111.10 (May 2025), page 103504. doi: 10.1103/PhysRevD.111.103504. arXiv: 2502.08707 [astro-ph.CO].
- [450] Keisuke Inomata. “Superhorizon Curvature Perturbations Are Protected against One-Loop Corrections”. In: 133.14 (Oct. 2024), page 141001. doi: 10.1103/PhysRevLett.133.141001. arXiv: 2403.04682 [astro-ph.CO].
- [451] Sachiko Kuroyanagi, Tomo Takahashi, and Shuichiro Yokoyama. “Blue-Tilted Tensor Spectrum and Thermal History of the Universe”. In: *Journal of Cosmology and Astroparticle Physics* 02.02 (Feb. 2015), page 003. doi: 10.1088/1475-7516/2015/02/003. arXiv: 1407.4785 [astro-ph.CO].
- [452] Sachiko Kuroyanagi, Tomo Takahashi, and Shuichiro Yokoyama. “Blue-Tilted Inflationary Tensor Spectrum and Reheating in the Light of NANOGrav Results”. In: *Journal of Cosmology and Astroparticle Physics* 01.01 (Jan. 2021), page 071. doi: 10.1088/1475-7516/2021/01/071. arXiv: 2011.03323 [astro-ph.CO].
- [453] Ken’ichi Saikawa and Satoshi Shirai. “Primordial Gravitational Waves, Precisely: The Role of Thermodynamics in the Standard Model”. In: *JCAP* 05.05 (May 2018), page 035. doi: 10.1088/1475-7516/2018/05/035. arXiv: 1803.01038 [hep-ph].
- [454] Ken’ichi Saikawa and Satoshi Shirai. “Precise WIMP Dark Matter Abundance and Standard Model Thermodynamics”. In: *JCAP* 08.08 (Aug. 2020), page 011. doi: 10.1088/1475-7516/2020/08/011. arXiv: 2005.03544 [hep-ph].
- [455] Bruce Allen. *The Stochastic Gravity-Wave Background: Sources and Detection*. Sept. 1996. doi: 10.48550/arXiv.gr-qc/9604033. arXiv: gr-qc/9604033.
- [456] Bruce Allen and Joseph D. Romano. “Detecting a Stochastic Background of Gravitational Radiation: Signal Processing Strategies and Sensitivities”. In: (1997). doi: 10.48550/ARXIV.GR-QC/9710117.
- [457] Michele Maggiore. “Gravitational Wave Experiments and Early Universe Cosmology”. In: *Physics Reports* 331.6 (July 2000), pages 283–367. doi: 10.1016/S0370-1573(99)00102-7. arXiv: gr-qc/9909001.

- [458] M Punturo et al. “The Einstein Telescope: A Third-Generation Gravitational Wave Observatory”. In: *Classical and Quantum Gravity* 27.19 (Oct. 2010), page 194002. doi: 10.1088/0264-9381/27/19/194002.
- [459] David Reitze et al. “Cosmic Explorer: The U.S. Contribution to Gravitational-Wave Astronomy beyond LIGO”. In: (2019). doi: 10.48550/ARXIV.1907.04833.
- [460] John Baker et al. “The Laser Interferometer Space Antenna: Unveiling the Millihertz Gravitational Wave Sky”. In: (July 2019). doi: 10.48550/arXiv.1907.06482. arXiv: 1907.06482 [astro-ph.IM].
- [461] Kai Schmitz. “New Sensitivity Curves for Gravitational-Wave Signals from Cosmological Phase Transitions”. In: *Journal of High Energy Physics* 2021.1 (Jan. 2021), page 97. doi: 10.1007/JHEP01(2021)097. arXiv: 2002.04615 [hep-ph].
- [462] Kai Schmitz. *New Sensitivity Curves for Gravitational-Wave Experiments*. Version v1. Feb. 2020. doi: 10.5281/zenodo.3689582. URL: <https://doi.org/10.5281/zenodo.3689582>.
- [463] LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration. *GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run — Data behind the figures*. [Data set]. 2021. doi: 10.5281/zenodo.5571767. URL: <https://doi.org/10.5281/zenodo.5571767>.
- [464] Gabriella Agazie et al. “The NANOGrav 15-Year Data Set: Detector Characterization and Noise Budget”. In: *The Astrophysical Journal Letters* 951.1 (July 2023), page L10. doi: 10.3847/2041-8213/acda88. arXiv: 2306.16218 [astro-ph].
- [465] Joseph Glaser et al. *Noise Spectra and Stochastic Background Sensitivity Curve for the NG15-year Dataset (1.0.0) [Data set]*. Version v1. June 2023. doi: 10.5281/zenodo.8092346. URL: <https://doi.org/10.5281/zenodo.8092346>.
- [466] Eric Thrane and Joseph D. Romano. “Sensitivity Curves for Searches for Gravitational-Wave Backgrounds”. In: *Phys. Rev. D* 88.12 (Dec. 2013), page 124032. doi: 10.1103/PhysRevD.88.124032. arXiv: 1310.5300 [astro-ph.IM].
- [467] Marco Drewes et al. “Towards a Precision Calculation of N_{eff} in the Standard Model. Part III. Improved Estimate of NLO Contributions to the Collision Integral”. In: *JCAP* 06.06 (June 2024), page 032. doi: 10.1088/1475-7516/2024/06/032. arXiv: 2402.18481 [hep-ph].
- [468] Tsung-Han Yeh, Keith A. Olive, and Brian D. Fields. “The Impact of New $d(p,\gamma)^3$ Rates on Big Bang Nucleosynthesis”. In: *JCAP* 03.03 (Mar. 2021), page 046. doi: 10.1088/1475-7516/2021/03/046. arXiv: 2011.13874 [astro-ph.CO].
- [469] O. Pisanti et al. “Primordial Deuterium after LUNA: Concordances and Error Budget”. In: *JCAP* 04.04 (Apr. 2021), page 020. doi: 10.1088/1475-7516/2021/04/020. arXiv: 2011.11537 [astro-ph.CO].
- [470] John T. Giblin and Eric Thrane. “Estimates of Maximum Energy Density of Cosmological Gravitational-Wave Backgrounds”. In: *Physical Review D* 90.10 (Nov. 2014), page 107502. doi: 10.1103/PhysRevD.90.107502.
- [471] Ogan Özsoy. “Synthetic Gravitational Waves from a Rolling Axion Monodromy”. In: 04 (Apr. 2021), page 040. doi: 10.1088/1475-7516/2021/04/040. arXiv: 2005.10280 [astro-ph.CO].

- [472] Caner Unal, Alexandros Papageorgiou, and Ippei Obata. “Axion-Gauge Dynamics During Inflation as the Origin of Pulsar Timing Array Signals and Primordial Black Holes”. In: *Physics Letters B* 856 (July 2024), page 138873. doi: 10.1016/j.physletb.2024.138873. arXiv: 2307.02322 [astro-ph.CO].
- [473] Milton Abramowitz and Irene A. Stegun, editors. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Ninth Printing. Washington, DC, USA: U.S. Government Printing Office, 1964.