A University of Sussex PhD thesis

Available online via Sussex Research Online:

http://sro.sussex.ac.uk/

This thesis is protected by copyright which belongs to the author.

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the Author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the Author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Please visit Sussex Research Online for more information and further details
Multiple field inflation and primordial non-Gaussianity

Kareem Marzouk

Submitted for the degree of Doctor of Philosophy

Department of Physics and Astronomy

University of Sussex

June 2022
Declaration

I hereby declare that this thesis has not been and will not be submitted in whole or in part to another University for the award of any other degree.

Original work:

- **Chapter 2** is drawn from the JCAP publication “Non-Gaussianity in D3-brane inflation”, by Marzouk, Maraio and Seery [1]. My key contributions to this paper were the PyTransport technology developments and sample analysis, in addition to theoretical work in understanding and constructing the inflationary model itself.

- **Chapter 3** is drawn from the JCAP publication “Constraints on $\tau_{\text{NL}}$ from Planck temperature and polarization”, by Marzouk, Lewis and Carron [2]. I was responsible for all of the numerical and data analysis, derived using codes developed by Carron. I additionally developed the spectral corrections method presented in the accompanying appendix.

Signature:

Kareem Marzouk
Summary: Multiple field inflation and primordial non-Gaussianity

This thesis is based on research into primordial non-Gaussianity produced during inflation. It focuses on predictions of the bispectrum from D3-brane inflation, and constraints on the squeezed-trispectrum derived from observational data.

Introduction

In Chapter 1, we introduce cosmology in a general context, outlining key successes of ΛCDM. We review homogeneous and perturbed cosmology, and identify the Cosmic Microwave Background (CMB) as a key observable that is referred to throughout this thesis. Subsequently, we introduce inflation as a theoretical framework to source the initial curvature perturbations. Particular emphasis is placed on multiple field inflation, which gives rise to non-Gaussian statistics that are later studied in depth. This chapter is closed with an interlude outlining the context for the original research in the following two chapters:

Non-Gaussianity in D3-brane inflation

In Chapter 2 we assemble numerical tools to study non-Gaussianities produced in multifield models of inflation. This advances the ‘Transport’ codes to samplers, that enable the exploration of parameter and observable spaces, and the correlations thereof. We explicitly apply these samplers to the D3-brane inflation scenario and provide the first results in the literature for the reduced bispectrum in this model. We attempt to provide a comprehensive and self contained account of the model itself, to assist future studies of the model.

Constraints on $\tau_{\rm NL}$ from Planck temperature and polarization data

In Chapter 3 we utilise the final data and simulations produced by the Planck collaboration to constrain the amplitude of the primordial trispectrum. Using temperature and polarization fields, we construct quadratic estimates of an underlying “modulation field” that can be translated into a constraint on $\tau_{\rm NL}$ via likelihood analysis. We provide the most stringent constraint in the literature to date, with a constraint $\sim 1.7 \times$ better than previously reported.

Closing remarks

In Chapter 4 we remark on the work produced in this thesis. We provide approximate relations between the topics of Chapter 2 and 3, before commenting on the future prospects of ‘Multiple field inflation and primordial non-Gaussianity.’
Acknowledgements

First and foremost, I would like to thank my PhD advisors Antony Lewis and David Seery. For the past years you have provided a huge amount of encouragement, support and mentoring, for which I am incredibly grateful. It is impossible to gauge how much I have taken away from working with you both, and I will certainly miss our discussions on ‘multiple field inflation and primordial non-Gaussianity’; as well many other things cosmology related.

Having been at Sussex for 9 years - from foundation year through to doctoral studies - there are many faculty members and friends who have all contributed to it being an excellent chapter of my life. I would like thank Chris Byrnes for many interesting discussions and advice throughout the course of my PhD, and to Mark Hindmarsh who sparked my interest in cosmology research whilst I was an undergraduate.

There are many PhD and postdocs that I have had the pleasure of discussing cosmology (often over a beer with after work on a Friday): Luke, Jussi, Maria, Chris, Pippa, Aswin, Steve, Raf, Lucas, Sean, Dan, Andrew, Giulio, Itzi, David, Will, et al. (I am sorry to have failed to mention you all!) From outside of astronomy there are many more who have entertained my love (and at times despair) for the work presented in this thesis, to name a few, Chris, Connor, Paul, Shaun, Alex, Fils, James, Noha, thank you! I would also like to thank Olivia for her immeasurable support throughout the past years - I am finally no longer a student, huzzah!

Finally, in addition to those at Sussex, I would like to thank the cosmology group at the University of Geneva for being so welcoming whilst I visited. In particular, thank you Julien for supporting my visit (as well as tolerating too many of my annoying questions over the years!), as well as Louis, Sebastian, Isabela and Lazare for keeping me excellent company.
# Contents

1 Introduction .............................. 1
   1.1 Elements of standard cosmology .................. 3
      1.1.1 FLRW spacetimes .......................... 3
      1.1.2 Cosmological perturbations ................... 7
      1.1.3 Gauge invariant scalars ...................... 10
      1.1.4 Evolution from initial conditions ............. 11
      1.1.5 Transfer functions .......................... 13
   1.2 Initial conditions from inflation .................. 16
      1.2.1 Single field inflation ......................... 18
      1.2.2 Remarks on Gaussian random fields .............. 23
      1.2.3 The $\delta N$ formalism ...................... 24
      1.2.4 Multiple field inflation ....................... 27
      1.2.5 The ‘Transport’ approach ...................... 31
   1.3 Sampling with PyTransport ....................... 36
   1.4 Context for original research ....................... 37
      1.4.1 Probing high-energy theory ..................... 37
      1.4.2 Constraining primordial non-Gaussianity .......... 39

2 Non-Gaussianity in D3-brane inflation ............... 42
   2.1 The D3/D3 model of inflation ...................... 53
      2.1.1 The conifold geometry ........................ 53
      2.1.2 The D3-brane potential ........................ 56
      2.1.3 Harmonic analysis on the conifold ............... 58
      2.1.4 Flux contributions ............................ 64
   2.2 Experimental procedure ............................ 71
      2.2.1 Software stack and sampling strategy .......... 71
      2.2.2 The adiabatic limit ........................... 80
      2.2.3 Agreement between pipelines .................... 82
   2.3 Results ........................................ 85
      2.3.1 Background evolution and mass spectrum ......... 86
      2.3.2 Two- and three-point observables ............... 91
      2.3.3 ‘Small’ ensemble comparisons .................... 107
   2.4 Discussion ...................................... 108

3 Constraints on $\tau_{\text{NL}}$ from Planck temperature and polarization 114
Chapter 1

Introduction

Cosmic inflation is the most developed theory of the Early Universe: A phase of rapid expansion prior to the onset of a hot big bang cosmology. It is fabled for its solution to the classical horizon, flatness and monopole problems in the otherwise standard cosmological picture [3, 4]. It has retained favour for its ability to source the initial seeds of large-scale structure, and the promises to one day be accounted for within a fundamental framework of high-energy physics.

Most models of inflation are parametrised by the evolution of one or more scalar fields. These fields may be fundamental, or an effective representation of some other microphysical quantity. Irrespective of their precise definition, under the canonical quantisation procedure of promoting classical fields to quantum fields [5], inflation leads to fluctuations on an expanding background. In the event that fluctuations grow to a size comparable to the horizon, then they transition from quantum to classical objects, perturbing the evolution of the Universe at later times [6].

Inflation ends when its constituent fields decay into standard model particles, flooding the Universe with quarks and gluons, and readying the onset of a hot big bang era. Throughout inflation the Universe becomes very cold due to the rapid dilution of energy content across an expanding cosmological volume. Therefore, the time when inflationary fields decay and deposit their energy into the Universe is referred to as reheating [7, 8, 9, 10, 11, 12]. Much like inflation itself, the details of reheating depend on underlying physical details. Most generically, particle production can be driven as fields oscillate around their minima, though there are other mechanisms such as annihilation; which can arise in string theoretic scenarios such as brane/anti-brane inflation [13, 14, 15]. The reheating phase is often approximated as instantaneous, though can be modelled more accurately as a perturbative process.

After reheating, the universe transitions through the short lived electroweak and quantum chromodynamic (QCD) epochs; giving rise to mediating bosons, and hadronizing the quark-gluon plasma [16, 17]. After $\mathcal{O}(10)$ seconds have elapsed, primordial nucleosynthesis (a.k.a. big bang nucleosynthesis, BBN) has further produced an abundance of light Helium nuclei [18] and traces of slightly heavier Lithium and Beryllium. If appro-
appropriate matter content was generated during reheating, primordial nucleosynthesis can be explained with standard nuclear particle physics, and is therefore a process that can be understood with terrestrial nuclear experimentation and constraints on the relic Helium abundances throughout the Universe today. In the near future it may be possible to constrain some details of the electroweak era with interferometer experiments such as LISA, due to the gravitational wave spectrum produced by its phase transition [19].

Following nucleosynthesis, the Universe embarks on its thermal history: evolving through states of radiation and matter domination, before that of accelerating expansion we observe today. Observations can be explained with remarkable consistency under the posited existence of a dark sector. This sector is comprised of some unknown “dark matter” [20, 21], which interacts gravitationally to assist the assembly of large-scale structure, and a form of “dark energy”, which is required to explain the accelerating expansion today [22]. The former of these components in part may be explained by compact relics from the early universe, such as primordial black holes [23, 24], but is widely believed to be a feature of particle physics beyond the standard model (see e.g. Kolb and Turner [25]). Dark energy could similarly be explained by extended physical models, though has a competing geometric interpretation as a “cosmological constant”; which has a direct correspondence with the Einstein field equations.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>68% limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Omega_b h^2$</td>
<td>0.02242 ± 0.00014</td>
</tr>
<tr>
<td>$\Omega_c h^2$</td>
<td>0.11933 ± 0.00091</td>
</tr>
<tr>
<td>$100\theta_{MC}$</td>
<td>1.04101 ± 0.00029</td>
</tr>
<tr>
<td>$\tau$</td>
<td>0.0561 ± 0.0071</td>
</tr>
<tr>
<td>$\ln(10^{10} A_s)$</td>
<td>3.047 ± 0.014</td>
</tr>
<tr>
<td>$n_s$</td>
<td>0.9665 ± 0.0038</td>
</tr>
</tbody>
</table>

Table 1.1: Base $\Lambda$CDM cosmological parameters inferred by the Planck collaboration 2018. Top to bottom: baryonic density parameter, dark matter density parameter, angular acoustic scale at recombination, optical depth to reionisation, amplitude of and tilt of primordial scalar perturbations. These references to ‘density parameter’ include a scaling by the dimensionless Hubble parameter

$$ h = \frac{H_0}{100\text{km s}^{-1}\text{Mpc}^{-1}}, $$

where $H_0$ is the Hubble rate of expansion today.

Collectively, these ingredients along with assumption on spatial curvature\(^1\) define the cosmological concordance model, for which, key observational predictors are founded in the cosmic microwave background (CMB). The CMB formed at the end of the radiation era when the cosmological fluid became sufficiently cool that light neutral atoms could assemble, and photons were relinquished from Thomson scattering via free electrons to freely travel the Universe. The relic free photons arising from this (re)combination process are red-shifted over cosmic times, becoming the primordial microwaves that we eventually

\(^1\)See following discussion in section 1.1.1.
observe today. In the concordance model, the CMB must therefore contain information pertaining to the early and late Universe; which can coarsely divided into pre- and post-recombination histories.

Through a spectral analysis of CMB observational maps, it is possible to constrain the relative composition of the Universe, since each cosmic component alters the spectral structure in distinct, albeit sometimes degenerate ways. Such an analysis is typically Bayesian in nature, since we are limited by the fact that we have only one observable Universe that we can get out and measure. The state-of-the-art analysis performed by the Planck collaboration \cite{2016A&A...594A..13P}\textsuperscript{2} suggests a best fit cosmological model which can be summarised by 6 base parameters in Tab. 1.1, with all additional parameters derived thereof. These core parameters describe the baryonic and cold dark matter density parameters of the Universe today, $\Omega_b$ and $\Omega_c$. The approximate angular acoustic scale at recombination, $\theta_{MC}$, defining a characteristic scale of acoustic waves in the primordial radiation fluid. The optical depth from reionisation, $\tau$, representing the information loss due to the (re-)scattering of free-streaming CMB photons off of the re-ionised inter-stellar medium generated by the first stars. Finally, the amplitude $A_s$ and tilt $n_s$ of the primordial scalar perturbations, summarising the initial conditions for cosmic evolution. The Planck data appear to be consistent with these initial perturbations being of superhorizon origin, highly close to Gaussian, and adiabatic \cite{2005PhRvD..72f3515L}. The adiabatic property is consistent with perturbations being sourced by a single effective degree of freedom.

### 1.1 Elements of standard cosmology

#### Conventions

In this discussion we adhere to a mostly positive metric signature $(-,+,+,+)$, with a choice of natural units such that $c = \hbar = 1$. We make use of the Einstein summation convention, where summation is implied over repeated indices. When using Greek indices the summation runs over all components of the metric, and when using Latin indices that sum is restricted to the transverse components only. To distinguish derivatives with respect to conformal and cosmic time, we will use $\prime \equiv d/d\tau$ and $\cdot \equiv d/dt$ respectively.

#### 1.1.1 FLRW spacetimes

Over the large scales of the Universe, when considering scales greater than $\sim 100\text{ Mpc}$, its content is observed to be statistically homogeneous and isotropic. These large scale properties can be rephrased as conditions of rotational and translational invariance, which

\textsuperscript{2}Here we are quoting the 68% confidence intervals, for TT,TE,EE+lowE+lensing+BAO.
in terms of a metric tensor can be written

\[ ds^2 = g_{\mu\nu} dX^\mu dX^\nu = -dt^2 + a^2(t) dx^2, \]  

(1.1)

where \( dt \) denotes an infinitesimal measure of time, and \( dx^2 = \gamma_{ij} dx^i dx^j \) characterises infinitesimal spatial distances. The transverse component of the metric \( \gamma_{ij} \) is controlled by spatial curvature, which may correspond to spherical, euclidean or hyperbolic spacetimes. These spacetimes are parametrized by the (comoving) 3-curvature

\[ K = \begin{cases} +1 & \leftrightarrow \text{Spherical} \\ 0 & \leftrightarrow \text{Euclidean} \\ -1 & \leftrightarrow \text{Hyperbolic} \end{cases}, \]  

(1.2)

with which we define the Friedmann-Lemaître-Robertson-Walker (FLRW) metric

\[ ds^2 = -dt^2 + a^2(t) \left( \frac{dr^2}{1 - Kr^2} + r^2 d\Omega^2 \right). \]  

(1.3)

We have chosen to represent the transverse coordinates in a spherical basis, such that \( r \) is radial distance (from an arbitrary center) and \( d\Omega^2 = d\theta^2 + \sin^2 \theta d\phi^2 \) is the associated solid angle. There is nothing special about this coordinate choice, but it enables the spatial curvature to be straightforwardly defined within the metric for a range of values without changing the functional form.

After assigning a spatial curvature, the only free function in the FLRW metric is the scale factor, \( a(t) \). In an expanding Universe this represents the relative size of spatial hypersurfaces at different times. Moreover, it is used to define the Hubble rate of expansion

\[ H = \frac{\dot{a}}{a}, \]  

(1.4)

which is positive in an expanding Universe or negative if contracting.

It is often convenient to consider conformal time \( \tau \), rather than cosmic time \( t \), which are related via the integral

\[ \tau = \int \frac{dt}{a(t)}. \]  

(1.5)

Conformal time directly relates to the comoving distance a photon may have travelled over cosmic time, and therefore serves as a type of horizon. With conformal time, the FLRW metric can be rewritten as

\[ ds^2 = a(\tau)^2 \left( -d\tau^2 + \frac{dr^2}{1 - Kr^2} + r^2 d\Omega^2 \right), \]  

(1.6)

such that the scale factor now appears as a conformal factor to the metric. Conformal time is useful since it directly relates time scales to distance scales via the scale factor. In
conformal time, we can also define the conformal Hubble rate

$$H = \frac{a'}{a} = aH.$$  \hspace{1cm} (1.7)

When performing calculations in curved (FLRW) spacetimes, we must extend the concept of derivative to covariant derivative. This enables us to preserve the properties of tensor transformation whilst moving through coordinate frames of curved spaces. The covariant derivative is computed as the partial derivative supplemented with a set of coefficients that linearly correct the change in coordinates. For vectors with covariant and contravariant indices these corrections are

$$\nabla_\omega T^\alpha = \partial_\omega T^\alpha + T^\delta \Gamma^\alpha_{\delta \omega}, \quad \nabla_\omega T_\alpha = \partial_\omega T_\alpha - T_\delta \Gamma^\delta_{\alpha \omega}. \hspace{1cm} (1.8)$$

The coefficients $\Gamma^\alpha_{\beta \gamma}$ represent the affine connexion, often referred to as “Christoffel symbols” (of the second kind). For a choice of metric $g_{\mu \nu}$ these are defined as

$$\Gamma^\alpha_{\beta \gamma} = \frac{1}{2} g^{\alpha \delta} (\partial_\beta g_{\delta \gamma} + \partial_\gamma g_{\beta \delta} - \partial_\delta g_{\beta \gamma}). \hspace{1cm} (1.9)$$

For higher rank tensor objects, the covariant derivative is computed analogously to the examples in Eq. (1.8), with the leading term defined as a partial derivative and each covariant index being linearly corrected with an affine connexion, e.g. $\nabla_\omega T_\beta^\alpha = \partial_\omega T_\beta^\alpha + T_\delta^\beta \Gamma^\delta_{\beta \omega} - T_\delta^\alpha \Gamma^\delta_{\beta \omega}$.

These definitions are important to understand the kinematics of particles in FLRW spacetimes. In particular, via the 4-velocity $u^\mu = dx^\mu/ds$, the equation of motion of a particle is given by the geodesic equation

$$\frac{du^\alpha}{ds} + \Gamma^\alpha_{\beta \gamma} u^\beta u^\gamma = u^\alpha \nabla_\alpha u^\mu = 0, \hspace{1cm} (1.10)$$

which in terms of 4-momentum reads

$$p^\alpha \partial_\alpha p^\mu = -\Gamma^\mu_{\alpha \beta} p^\alpha p^\beta. \hspace{1cm} (1.11)$$

The energy and mass of the particle are then related to the momentum via $p^2 = a^2 \gamma_{ij} p^i p^j = E^2 - m^2$, which for a particle moving exclusively along the radial direction of an FRLW spacetime where spatial derivatives of the 4-momentum vanish, Eq. (1.11) yields the important result

$$\frac{\dot{p}}{p} = -\frac{\dot{a}}{a}. \hspace{1cm} (1.12)$$

For particles moving along geodesics, their momentum is inversely proportional to the rate of expansion: $p(t) \propto a(t)^{-1}$. This means that the 3-momentum of a massive particle decays, and the wavelength of photons is stretched as a Universe expands (since their wavelength is inversely proportional to their energy). This process is known as ‘red shift’, and is central to many cosmological inferences.
The FLRW metric provides an appropriate framework for characterising cosmological distances and symmetries. On its own, however, it does not describe the matter and energy content of the Universe. To do this, we additionally require the Einstein equations

\[ G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = 8\pi T_{\mu\nu} + \Lambda g_{\mu\nu}. \] (1.13)

Here the Einstein tensor \( G_{\mu\nu} \) is defined in terms of spacetime curvature, or equivalently, in terms of matter and energy. In the former case \( G_{\mu\nu} \) is constructed from the Ricci scalar \( R = g^{\mu\nu} R_{\mu\nu} \) and Ricci tensor \( R_{\mu\nu} = R^\alpha_{\mu\alpha\nu}, \) which are contractions of the Riemann tensor

\[ R^\alpha_{\beta\gamma\delta} = \partial_\gamma \Gamma^\alpha_{\delta\beta} - \partial_\delta \Gamma^\alpha_{\gamma\beta} + \Gamma^\omega_{\delta\beta} \Gamma^\alpha_{\gamma\omega} - \Gamma^\omega_{\gamma\beta} \Gamma^\alpha_{\delta\omega}. \] (1.14)

This highlights a detail we have previously omitted, that a FLRW spacetime must be (pseudo-)Riemannian manifold; as to ensure e.g. smoothness and differentiability. The matter and energy content enter on the right hand side, and are described by the energy-momentum (EM) tensor \( T_{\mu\nu} \) along with a cosmological constant, \( \Lambda. \)

Direct computation from the FLRW metric yields the non-zero components of the Einstein tensor

\[ G_{00} = 3 \left[ \frac{\dot{a}}{a} \right]^2 + \frac{K}{a^2}, \quad G_{ij} = - \left[ 2 \frac{\dot{a}}{a} + \left( \frac{\dot{a}}{a} \right)^2 + \frac{K}{a^2} \right] g_{ij}. \] (1.15)

This tensor is diagonal, as expected by the underlying isotropy and homogeneity of FLRW spacetimes. For self consistency, the Einstein equations require that \( T_{\mu\nu} \) must also be diagonal. Homogeneity implies that the transverse components must be equal, and moreover, since the FLRW metric is time-dependent, so should the EM tensor. The functional form of the EM tensor is therefore restricted to

\[ T^\mu_\nu = \text{diag}(-\rho(t), P(t), P(t), P(t)), \] (1.16)

where \( \rho \) and \( P \) define the energy density and pressure of the cosmological fluid. In general, the total fluid may be a mixture of species, ‘\( a \)’, which collectively define the overall density and pressure in the sense that \( \rho = \sum_a \rho^{(a)} \) and \( P = \sum_a P^{(a)}. \) There are usually multiple species present, though for a given time, the Universe is typically dominated at the background level by a specific type, e.g. matter, radiation, or \( \Lambda. \) This means that in a multi-component cosmology, the evolution of the net fluctuations are dominated by those of the species that dominates the background. Since the EM tensor represents conserved quantities, it must satisfy \( \nabla_\mu T^\mu_\nu = 0. \) This gives rise to the continuity equation

\[ \dot{\rho} + 3H(\rho + P) = 0, \] (1.17)

which is recognisable from traditional fluid mechanics, though now accounts for the dilution of matter and energy through expansion. The factor of 3 in Eq. (1.17) follows from tracing over the the 3-spatial components of the EM tensor, and it holds for the sum over and individual particle species simultaneously.
The energy and pressure of a fluid specie can often be related via a linear equation of state $P^{(a)} = wP^{(a)}$. If so, this means Eq. (1.17) has a simple analytic solution

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

(1.18)

There are three canonical cases of $w$ that are considered significant throughout cosmic history

$$w = \begin{cases} 
0 & \text{Pressureless matter: e.g. cold dark matter and baryons} \\
1/3 & \text{Radiation: e.g. bosons and neutrinos} \\
-1 & \text{Vacuum energy: e.g. } \Lambda \text{ or dark energy}
\end{cases}.$$  

(1.19)

The “importance” stems from the fact that when a given species dominates the cosmological background, their equation of states determines the rate of expansion. In particular, $\rho \propto a^{-3}$ when dominated by pressureless matter (a.k.a. “dust”), $\rho \propto a^{-4}$ when dominated by radiation, and $\rho \propto \text{constant}$ when dominated by some form of vacuum energy. Alternatively these statements may be written in terms of cosmic time $t$, such that the equation of state governs the time evolution of the expansion at a given epoch.

By combining the curvature and EM equivalences of the Einstein tensor, the Friedmann equations follow:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3} \rho - \frac{K}{a^2}, \quad \frac{\ddot{a}}{a} = -\frac{4\pi G}{3} (\rho + 3P).$$  

(1.20)

The first Friedmann equation can be used to define the critical density, i.e. the density required to produce a flat Universe with $K = 0$

$$\rho_c = \frac{3}{8\pi G} \left(\frac{\dot{a}}{a}\right)^2.$$  

(1.21)

The critical density is useful to benchmark the energy budget of the Universe in way that enables theoretical predictions to be compared with data. In particular, we define the cosmological density parameter of a species $a$ at time $t_*$

$$\Omega_a|_{t=t_*} = \frac{\rho_a}{\rho_c|_{t=t_*}}.$$  

(1.22)

Denoting the age of the Universe today as by $t_* = t_0$, the sum over all density parameters satisfies $\sum_a \Omega_a|_{t=t_0} = 1$ if the Universe is flat; $1 < \sum_a \Omega_a|_{t=t_0}$ if hyperbolic, or $\sum_a \Omega_a|_{t=t_0} > 1$ if positively curved.

1.1.2 Cosmological perturbations

The Universe is not perfectly homogeneous and isotropic. We can see this in the early Universe from the anisotropies of the microwave background at the $\sim 10^{-5}$ level, and in the later Universe from the clustering of matter into galaxies and other large structures.
In order to characterise these deviations from the homogeneous background, the starting point is linear cosmological perturbation theory. In essence, this provides a framework to dynamically evolve the perturbations that arise in the radiation era as they fall into contact from superhorizon scales. This relies on constructing a metric and accompanying EM tensor, that permit deviations from the homogeneous background in the form of scalar, vector and tensor (SVT) perturbations.

Starting with the metric, we may generalise Eq. (1.6) to

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

(1.23)

where \( A \) is built from a pure scalar, \( B_i = B_i(B,B,V) \) from scalars and vectors, and \( h_{ij} = h_{ij}(C,E,E^V,E_{ij}) \) from scalars, vectors and tensors. Collectively, there are an additional 10 degrees of freedom; that are introduced in this so-called “SVT decomposition”. In the absence of metric vector and tensor perturbations, the FLRW metric can be written with generality as

\[ ds^2 = a(\tau)^2 \left( -(1 + A)d\tau^2 + 2\partial_i B dx^i d\tau \right. \]

\[ \left. + \left[ (1 + 2C)\delta_{ij} + 2 \left( \partial_i \partial_j - \frac{1}{3} \delta_{ij} \nabla^2 \right) E \right] dx^i dx^j \right). \]

(1.24)

These additional degrees of freedom were added to the metric “by hand” to accommodate general perturbations to the background. However, we could have obtained a similarly “perturbed” metric structure under coordinate transformations, without intentionally altering the structure of the metric in any particular way. For instance, if \( x^i \to x^i + \xi^i \) and \( \tau \to \tau + \xi^0 \), then the background metric acquires terms that mimic scalar and vector perturbations. This effect is un-physical, and the degrees of freedom obtained in this way are known as “gauge” modes. This leads to an ambiguous interpretation of the perturbed metric, but fortunately, gauge modes can be removed via gauge fixing.

This process amounts to setting quantities prescribed via the SVT decomposition to zero. Common gauge choices include the synchronous, spatially flat and Newtonian gauge, in which \( A = B = 0, C = E = 0 \) and \( B = E = 0 \) respectively. Working in the synchronous gauge, is akin to viewing hypersurfaces with constant time. Without tensors, the spatially flat gauge leaves transverse components of the metric unperturbed. In the Newtonian gauge, spatial perturbations can be thought of as Newtonian gravitational potentials. Another option beyond gauge fixing is to work with gauge invariant quantities. These were famously introduced by Bardeen [28], in the form of the potentials

\[ \Psi = A + \mathcal{H}(B - E') + (B - E')', \quad \Phi = -C - \mathcal{H}(B - E') + \frac{1}{3} \nabla^2 E. \]

(1.25)

These methods of dealing with gauge modes are not mutually exclusive. In particular, by fixing spatially flat and Newtonian gauges, whilst making use of the Bardeen potentials,
we obtain the perturbed Conformal Newtonian Gauge (CNG) metric

\[ ds^2 = a(\tau)^2 \left[ -(1 + 2\Psi)d\tau^2 + (1 - 2\Phi)dx^2 \right]. \]  

(1.26)

This choice of metric is popular since its physical interpretation is clear: We can concisely represent any perturbation along temporal or spatial metric components in terms of scalar potentials.

Having settled on a suitable metric, the EM tensor assumes the form \( T_{\mu\nu} = \bar{T}_{\mu\nu} + \delta T_{\mu\nu} \). Small perturbations have simply been added to the homogeneous background quantities which are now denoted with an over-bar. When the cosmological fluid is perturbed, the EM tensor can receive non-diagonal corrections. This occurs via local bulk velocity flows \( v^i \) which previously did not exist, but are now sourced by spatial gradients arising from perturbations. Components of the EM tensor now take the form

\[
\begin{align*}
T^0_0 &= \bar{\rho} + \delta \rho, \\
T^i_0 &= (\bar{\rho} + \bar{P})v^i, \\
T^i_j &= -(\bar{P} + \delta P)\delta^i_j + \Pi^i_j.
\end{align*}
\]

(1.27, 1.28, 1.29)

Compared to the previous background EM tensor, there is a new factor of “anisotropic stress” \( \Pi^i_j \). This is non-zero when spatial and temporal potentials are not aligned, i.e. when \( \Delta \Phi \Psi \equiv \Phi - \Psi \neq 0 \) in the CNG. For late Universe standard cosmologies it is typically approximated that \( \Delta \Phi \Psi = 0 \), though does not hold exactly for a more accurate treatment of neutrinos [29]; and in more complex models of dark energy [30]. At linear order, conservation of the EM tensor (\( \nabla_\mu T^{0\mu} = 0 \)) implies two continuity equations - one at zeroth and one at first order in perturbations - which read

\[
\begin{align*}
\bar{\rho}' &= -3H(\bar{\rho} + \bar{P}), \\
\delta \rho' &= -3H(\delta \rho + \delta P) + (\bar{\rho} + \bar{P})(3\Phi' - \nabla \cdot v),
\end{align*}
\]

(1.30, 1.31)

where \( v \) is the 3-velocity vector of the fluid. Similar expressions can be derived for the relativistic Euler equation by conserving the transverse sector \( T^{i\mu} \).

As with the metric, we should consider EM perturbations in a gauge context, or construct them in invariant ways. For example, we can perform calculations in a gauge of uniform density (when \( \delta \rho = B = 0 \)), or in the comoving gauge corresponding to the rest frame of a fluid (when \( v = B = 0 \)). To understand the evolution of a given fluid component, it is convenient to define the gauge-invariant \textit{comoving density contrast}

\[
\Delta = \frac{\delta \rho}{\bar{\rho}} + \frac{\bar{\rho}'}{\bar{\rho}}(v + B),
\]

(1.32)

which in the comoving gauge is simply \( \delta = \delta \rho / \bar{\rho} \). It is helpful to define the fluid sound speed \( c_s^2 = \delta P / \delta \rho = P'/\rho' = \bar{P}/\bar{\rho} \), where the latter two equalities hold if the fluid satisfies
\( w = P/\rho \). We can then, for example, write the relativistic continuity equation as

\[
\delta' + (1 + w)(\nabla \cdot \mathbf{v} - 3\Phi') + 3\mathcal{H}(c_s^2 - w)\delta = 0.
\]

(1.33)

From this linearised analysis, solutions to the Einstein equations tell us about the dynamical evolution of the perturbations in terms of the metric potentials. In particular, we obtain the relativistic Poisson equation

\[
\nabla^2 \Phi = 4\pi G a^2 \bar{\rho} \delta,
\]

(1.34)

and the closed form expression for the evolution of the Potential itself,

\[
\Phi'' + 3(1 + w)\mathcal{H} \Phi' - w \nabla^2 \Phi = 0.
\]

(1.35)

Note that here we require \( w \) to be defined in terms of the mean fluid density and pressure.

1.1.3 Gauge invariant scalars

As we have discussed, FLRW spacetimes enable us to describe an isotropic and homogeneous cosmological background; in addition to one in which dynamical perturbations can grow on smaller scales. In inflationary cosmology, we particularly care about the initial superhorizon perturbations that seed inhomogeneities in the radiation fluid.

Before considering this explicitly, let us consider a perturbed cosmology more generally. For an arbitrary perturbation that we might care about, we should be able to find some local patch of the Universe that can be accurately modelled as homogeneous and isotropic, so long as the wavelength of that perturbation exceeds the size of the local patch. In this way, we can imagine a perturbed cosmology as a collection of smoothly connected separate Universes; each of which is essentially a smaller FLRW spacetime. Each patch should inherit a spatial curvature \( K \) from the global spacetime, but we know that each patch of the Universe will differ in its precise configuration of density and pressure perturbations.

Under a choice of gauge, whereby \( \delta \rho = 0 \), these local differences in perturbations are then shifted to local differences in curvature. We can characterise spatial curvature directly from the FLRW metric Eq. (1.3) with the Ricci 3-curvature \( R^{(3)} = 6K/a^2 \). In the comoving frame, this leads to the definition of the gauge invariant scalar quantity

\[
\mathcal{R} = \frac{a^2 \bar{R}^{(3)}}{4k^2}.
\]

(1.36)

which follows from \( 4\nabla^2 \mathcal{R} = -a^2 \bar{R}^{(3)} \), and the over-bar here denotes the local evaluation of curvatures.

Alternatively, we could write down an FLRW-like metric that absorbs local perturbations
directly into the transverse sector by introducing some quantity $\zeta$

$$ds^2 = -dt^2 + a(t)^2 e^{2\zeta(x,t)} dx^2. \quad (1.37)$$

This would be equivalent to perturbing the local scale-factor; $a(x, t) = a(t)(1 + \zeta(x, t))$ if we choose $\zeta = \delta \ln a$. Computing the Ricci scalar directly for this metric yield $a^2 R^{(3)} = -4\nabla^2 \zeta$; and therefore $\zeta$ is the same as $R$ when evaluated in the $\delta \rho = 0$ gauge. For this reason, $\zeta$ is sometimes known as the curvature perturbation of constant density. It can be written in gauge invariant ways

$$\zeta - H\frac{\delta \rho}{\rho} = \zeta + \frac{\delta \rho}{3(\rho + p)}, \quad (1.38)$$

and can be related to the comoving curvature perturbation via

$$R = \zeta + \frac{\delta \rho}{3(\rho + p)}. \quad (1.39)$$

In the conformal Newtonian gauge we can identify $\zeta = -\Phi$, and with some manipulation write the curvature perturbation in terms of the equation of state for the background

$$\zeta = \Phi - \frac{1}{3(1 + w)}\delta = \frac{5 + 3w}{3 + 3w} \Phi. \quad (1.40)$$

Note that in the divergence $\lim_{w \to -1} \zeta = \inf$ is non-physical, corresponding to the fully superhorizon limit which is an unobservable background redefinition. By comparison, the null contribution $\zeta|_{w=-5/3} = 0$ corresponds to ...

For adiabatic perturbations, it can be proven with generality that $\zeta$ is conserved across superhorizon scales, as shown by Lyth, Malik and Sasaki [31]. By association, it further follows that $R$ is also conserved on superhorizon scales. It is for this reason that these perturbations form an integral part of inflationary analyses. Quantum fluctuations generate field fluctuations on sub-horizon scales during an inflationary epoch, but since the Hubble volume is decreasing as time goes forwards, these fluctuations transcend to superhorizon scales.

In the separate Universe picture, these fluctuations are mapped to superhorizon curvature perturbations which are then conserved on superhorizon scales. Once inflation ends and the Hubble volume starts to expand, curvature perturbations will initial enter local horizons, and provide a source of initial conditions from which inhomogeneities can grow, as sketched in Fig. 1.1.

1.1.4 Evolution from initial conditions

As discussed, the conserved scalars can be related to the initial conditions for perturbations during radiation dominance. To see this, we note that on superhorizon scales, large enough such that spatial gradients are negligible, the closed form expression for the evolution the
Figure 1.1: Horizon exit and re-entry of comoving curvature perturbations. Figure taken directly from [32].

potential in Eq. (1.35) implies that $\Phi = \text{const}$. Recalling that $\zeta$ can be written in terms of the equation of state for the background, from Eq. (1.40) it follows that on super horizon scales

$$\zeta|_{\text{Rad. Dom.}} = \frac{3}{2}\Phi.$$  \hspace{1cm} (1.41)

Using Eq. (1.35) once again, during radiation dominance we can write the mode equation

$$\Phi'' + \frac{4}{\tau}\Phi' + \frac{k^2}{3}\Phi_k = 0,$$  \hspace{1cm} (1.42)

which admits solutions that are proportional to the initial curvature configuration $\Phi_k(\tau) \propto \zeta_k(0)$. The superhorizon solution is simply constant as per Eq. (1.41), and on subhorizon scales the solution is oscillatory

$$\Phi_k(\tau) \propto \zeta_k(0)\frac{\cos(k\tau/\sqrt{3})}{k^2\tau^2}.$$  \hspace{1cm} (1.43)

In order to understand what this means for the growth of perturbations, we recall that the Poisson equation relates the density contrast to the Newtonian potential. This means that the density contrast of the radiation scales like $\Delta_r \propto (k\tau)^2\Phi_k \propto a^2\Phi_k$; such that the evolution of the density contrast resembles the evolution of the Potential, multiplied by the scale factor squared.\(^3\) Physically, this corresponds to the red shifting of radiation over superhorizon scales. On subhorizon scales, the oscillating radiation corresponds to the competing gravitational in-fall of radiation and its expansive response to pressure, initialised by the potential before it decays away.

Whilst radiation is dominant, there are also sub-dominant baryonic and dark matter species. The baryonic content in the radiation era is a catch all for electrons and protons, the former of which will interact with the radiation photons through Thomson scattering.

\(^3\)Note that the $(k\tau)^2$ cancels on subhorizon scales.
This leads to the photons being tightly coupled with the baryons. Therefore, the evolution of the baryonic density contrast $\Delta_b$ traces the photons\(^4\) of the radiation $\Delta_r$ in such a way that the baryons oscillate with the photons. Meanwhile, dark matter does not couple to photons as the baryons do. Deep inside the horizon, however, dark matter will initially coalesce towards the gravitational minima before the potential decays away, with its density contrast evolving logarithmically in a way known as the Meszaros growth [33]. Over large scales, this means that the dark matter perturbations lead to an enhanced disparity between over- and under-dense regions in the coupled radiation fluid.

This identification of oscillatory solutions within the primordial photon-baryon fluid offers a qualitative insight as to how the CMB acoustic structure forms. Evidently, the radiation era does not last forever, ultimately ending with recombination once the Universe becomes sufficiently cooled by the work done under expansion. Therefore, at the time of recombination, photons will necessarily be at some arbitrary point of the acoustic cycle: the photons will be in a hot state of compression or cooler state of rarefaction. This information will map into the position space distribution of CMB photon temperatures as they begin to stream from the last-scattering surface. Before they can do so, however, they must escape from their respective potential wells at the expense of gravitational red shifting [34]. This gives rise to the primary CMB anisotropies from the time of recombination that we can observe today.

As the CMB photons begin to decouple, they will thermally diffuse about the recombination surface, and therefore wash information away as they propagate from hot to cold regions [35]. This effect decays as the Universe becomes more optically thin, since the particle content is diluted away. For this reason it is only the small scale information that is effected.

1.1.5 Transfer functions

There are many physical effects that can alter the properties of a CMB photon, from its dynamical evolution up to recombination, to its reprocessing as it travels across the Universe at later times. In terms of observables, however, what we really care about is the statistical properties of the distribution of photons as we observe them across the sky today. This is what we can measure, and what we can use to confront theoretical predictions with observations.

In order to do this, it is practical to consider the photon (phase space) distribution function, $f$, which in general depends on time, position and momentum. It is straightforward to see that a CMB photon in thermal equilibrium can be related to a temperature distribution. Neglecting chemical potentials, the equilibrium photons can be written as a blackbody distribution [36]

$$f(E) = \frac{1}{e^{E/T} - 1}, \quad (1.44)$$

\(^4\)When referring to radiation we are mostly referring to the photons, but there are other relativistic
which can be integrated to compute the number density $n$, and energy density $\rho$. The energy density of the radiation then satisfies a Stefan-Boltzmann relation of the form $\rho \propto g_{\text{eff}} T^4$, where $g_{\text{eff}}$ represents the effective number of degrees of freedom within the radiation fluid and $T$ is the mean background temperature. Each species of the total cosmological fluid will in fact contribute to the total energy density via $\rho = \sum_a \rho^{(a)}$; which relates fluid components by their equilibrium temperature.

The time evolution of the distribution function can be written in terms of some collision operator $C[f]$, which keeps track of interactions such as Thomson scattering, in the case of primordial photons. This defines the Boltzmann equation for a given distribution function

$$\frac{df}{dt} = C[f], \quad (1.45)$$

where $C$ in general could describe an arbitrary numbers of particle interactions. In order to model temperature anisotropies, we must however consider the perturbed form of Eq. (1.45). This requires perturbing the distribution function $f \to f + \delta f$, which in the case of CMB photons actually retains a black body spectrum. This means that perturbations in temperature are actually independent of the magnitude of the photon momentum.

This leads to the definition of the brightness function [36]

$$\Theta(\tau, x, \mathbf{n}) = \frac{\delta T(\tau, x, \mathbf{n})}{T(\tau)}, \quad (1.46)$$

where $\mathbf{n} = p/p$ is the direction of the photon momentum.

The brightness distribution can be expanded into a harmonic basis for a fixed time

$$\Theta(\tau, x, \mathbf{n}) = \sum_{\ell m} \Theta_{\ell m}(\tau, x) Y_{\ell m}(\mathbf{n}), \quad (1.47)$$

where $x$ represents the location of an observer, $Y_{\ell m}$ denotes the spherical harmonics, and $\Theta_{\ell m}(\tau, x)$ the set of coefficients that define the amplitude and phase of a given spherical mode defined by $\{\ell, m\}$. The $\ell = 0$, and $\ell = 1$ modes define the monopole and dipole contributions, the former of which is related to the photon density contrast, and the latter the relative motion of an observer.

With the assistance of the Legendre polynomials, the brightness function can be written in terms of the harmonics

$$\Theta_{\ell m}(\tau, x) = \frac{4\pi}{(2\pi)^3} \int \Theta_{\ell}(\tau, \mathbf{k}) Y_{\ell m}^{*}(\mathbf{k}') e^{i\mathbf{k} \cdot \mathbf{x}} d^3\mathbf{k}. \quad (1.48)$$

Choosing a fixed observer position $x = 0$ at a time $\tau = \tau_0$, we define the scalar function $\Theta_{\ell}(\mathbf{k}) = \Theta_{\ell}(\tau_0, \mathbf{k})$. The CMB multipoles observed today can then be written as

$$a_{\ell m} = \frac{4\pi}{(2\pi)^3} \int \Theta_{\ell}(\mathbf{k}) Y_{\ell m}^{*}(\mathbf{k}') d^3\mathbf{k} = \frac{4\pi}{(2\pi)^3} \int_0^{\infty} T_{\ell}(k) \xi_{\ell m}(k) \frac{dk}{k}. \quad (1.49)$$

components such as neutrinos, which do not couple in the same way.

5Write a bit about the transfer function here and insert some refs...
In this last equality we have introduced the transfer function $T_\ell(k)$ which tells us how to linearly evolve a given quantity subject to some initial primordial perturbation. In order to compute the harmonic coefficients, we need $\theta_\ell$. This data in fact comes from the perturbed Boltzmann equation, which define a hierarchy of differential equations, in which $\dot{\Theta}_{\ell+1}$ can be solved in terms of $\dot{\Theta}_\ell$.

In practice, there are now mature “Boltzmann codes” such as CAMB [37] and CLASS [38, 39], that can solve large hierarchies; with more complexities than the minimal photon-baryon interaction previously discussed. Moreover, these codes wrap additional recombination software packages such as RECFAST [40] and HyRec [41], that enable non-trivial recombination processes to be modelled with more detail than the simplistic Hydrogen assembly. The key output from executing these codes is a prediction for the CMB power spectrum. For the temperature fluctuations this is essentially defined in terms of the variance of the harmonics in Eq. (1.49)

$$\delta_\ell \delta_m C_\ell^\Theta = \langle a_\ell m a_\ell^* m' \rangle = \delta_\ell \delta_m P(\zeta)(k) \frac{dk}{k}, \tag{1.50}$$

where the amount of power in a given multipole is sensitive to the set of input parameters that define the cosmological model. Though differing by their implementation, these codes essentially compute a “source” term via solving the Boltzmann hierarchy up until some appropriate multipole truncation, before performing a line-of-sight integral in Fourier space [42, 43].

In Figure 1.2 we illustrate this process by computing the lensed CMB signal using CAMB. In this example, we choose to vary the ratio of dark and baryonic matter densities, defined in terms of their abundance in the late Universe today. The total matter content is kept constant, such that $\Omega_m = \Omega_b + \Omega_c$, and the remaining cosmological parameters are fixed. By varying the ratio $\Omega_b/\Omega_c$ we are effectively emulating the effect of loading more baryons into the primordial fluid of the early Universe at the expense of less dark matter. More baryons enables greater compression of the primordial fluid, in term extending the

![Figure 1.2: Theoretical lensed CMB temperature spectra, with varying ratio of baryonic to cold dark matter content.](image)
compressive phase of the acoustic oscillations. Greater compression over rarefaction leads to an asymmetric change in amplitude across the multipole peaks: the odd peaks being enhanced, and the even being suppressed. The reduced oscillation frequency from extended compression further results in an a shifted CMB spectrum towards smaller scales.\textsuperscript{6} Thus the matter content has a direct impact on the CMB acoustic structure, which can be seen by eye in the figure.

In addition to temperature, there is also a polarisation power spectrum. The primary source of polarisation from primordial photons arises from their Thomson scattering with free electrons. The incident photon acquires a linear polarization as it re-scatters outwardly, perpendicular to the ingoing direction. Multiple scattering events about a common electron can produce unpolarized, partially polarized, or completely polarized outgoing radiation depending on the geometry of incident rays. Since there are countless scattering events prior to recombination, there is a null polarising effect on average throughout this time. Only over the last scattering events do the CMB photons retain a distinct polarisation signature. During this time, any velocity gradients of the photon-baryon fluid arising from flow into and out from potential peaks and troughs Doppler shifts the photons. This Doppler shift induces a quadrupolar polarisation, radially as photons flow towards cold spots, and transversely as they move out form within wells \cite{44, 45}. The CMB polarisation anisotropy is therefore correlated with the temperature anisotropy generated at recombination by virtue of the velocity flow into and out from hot and cold spots.

\subsection*{1.2 Initial conditions from inflation}

As we have discussed, the evolution of perturbations that give rise to the CMB spectrum we observe today appear to be consistent with a set of initial conditions that propagated from superhorizon origin. We will now discuss how these perturbations could have been generated through inflation. In order to set the stage let us briefly review what is meant by the horizon problem in cosmology without inflation.

From observations of the CMB, we know that the early Universe was isotropic and homogeneous at the time of recombination, $\tau_{\text{rec}}$, characterised by a uniform background radiation temperature; and supplemented with small fluctuations at the $\sim 10^{-5}$ level. In a Universe that did not undergo inflation, this a conundrum, in the sense that within the conformal time that elapsed between $\tau = 0$ and $\tau = \tau_{\text{rec}}$, classical processes could have only correlated information across angular scales of up to $\sim 1^\circ$ as observed today. Explaining any statistical correlations beyond these scales is then problematic; unless the Universe had an incredibly precise initial configuration.

The inflationary solution to this problem is to introduce a phase of transient exponential expansion before the birth of the hot big bang era. This assumes that the Universe was causally connected before inflation began, and by exponential expansion, has the ability

\textsuperscript{6}A sideways shift can also be obtained when altering the sound horizon.
to place causal information beyond classically uncorrelated distances.

This is most easily understood visually as per Fig. 1.3. In this conformal time diagram, the maximum distance that information can be communicated is represented by the past light cone; which is defined by 45 degree lines. We can think of ourselves as an observers at $\tau_0$ looking backwards through time towards the recombination surface at $\tau_{\text{rec}}$. The orange patches between the recombination and the hot big bang (here denoted by “reheating”) represent the distances that information could be communicated over. Without overlap, it is implied that these patches could not have come into causal contact between $\tau = 0$ and $\tau_{\text{rec}}$. In effect, inflation introduces a span of negative conformal time that extends backwards from $\tau = 0$ towards some initial singularity. This permits information to propagate from some earlier causally connected region of the Universe.

In a model independent way, we can describe the process of inflation geometrically. During inflation, we require the Hubble volume to shrink, such that patches of the Universe can become causally disconnected. Via proportionality to the Hubble Horizon, this amounts to the condition

$$\frac{d}{dt}(aH)^{-1} < 0.$$  \hfill (1.51)

Direct computation reveals that this condition can be equivalently written as

$$\epsilon_H = -\frac{\dot{H}}{H^2} < 1;$$  \hfill (1.52)

such that the Hubble rate must be slowly varying over the course of inflation. In the
de Sitter limit whereby \( \epsilon_H \to 0 \) and \( H = \text{const.} \), the solutions for the scale factor are exponential, \( a(t) \propto e^{Ht} \). Assuming an FLRW background, the condition for inflation Eq. (1.52) can be written in terms of the pressure and energy density of an idealised fluid

\[
\epsilon_H = \frac{3}{2} \left[ 1 + \frac{P}{\rho} \right].
\]

We can imagine that the Universe was filled with a fluid satisfying \( P < -\rho/3 \); or in other words, having an equation of state \( w < -1/3 \).

### 1.2.1 Single field inflation

For the simplest models of inflation the fluid can be thought of as a single scalar field \( \phi \) minimally-coupled to gravity. The corresponding action for this field is [4, 46]

\[
S = \int d^4x \sqrt{-g} \left[ \frac{1}{2} R - \frac{1}{2} \partial_\mu \phi \partial^\mu \phi - V(\phi) \right],
\]

where \( R \) is the Ricci scalar and \( V(\phi) \) is the fields potential; describing its self interaction. Eq. (1.54) is simply the sum of Einstein-Hilbert and canonical scalar field actions.

Taking \( \phi \) to be homogeneous, then its spatial derivatives vanish, and it follows from the EM tensor that the energy and pressure of the field are given by

\[
\rho_\phi = \frac{1}{2} \phi^2 + V(\phi), \quad P_\phi = \frac{1}{2} \phi^2 - V(\phi).
\]

Evidently, the equation of state easily satisfies \( w < -1/3 \) if \( \phi^2 \ll V(\phi) \). Minimising Eq. (1.54) with respect to \( \phi \), and making use of the Friedmann equations leads to the background equations

\[
3H^2 = \frac{1}{2} \dot{\phi}^2 + V(\phi), \quad 0 = \ddot{\phi} + 3H \dot{\phi} + \partial_\phi V;
\]

such that time derivative of the Hubble rate is related to the kinetic energy of the field \( H = -\frac{1}{2} \dot{\phi}^2 \). In conformal time, Eqs. (1.56) can be written

\[
3\dot{H}^2 = \frac{1}{2} (\phi')^2 + a^2 V(\phi), \quad 0 = \phi'' + 2H \phi' + a^2 \partial_\phi V.
\]

By combining the background evolution equations, the condition for inflation in terms of the scalar scalar field

\[
\epsilon_H = \frac{1}{2} \frac{\phi^2}{H^2} < 1.
\]

This helps identify the key phenomenology of the scalar field that gives rise to inflation. For inflation to occur, we need \( \epsilon_H < 1 \) which corresponds to the potential exceeding the kinetic energy of the field; \( V(\phi) \gg \dot{\phi}^2/2 \). Moreover, we can advocate a prolonged period of inflation, by making the acceleration small in the sense that \( |\ddot{\phi}| \ll 3H |\dot{\phi}| \). Inflation characterised by these properties is what is known as slow-roll (SR), due to the schematic
picture of a scalar field slowly rolling down its potential. In the slow-roll approximation, it follows that Eqs. (1.56) become

\[ 3H^2 \approx V(\phi), \quad 0 \approx 3H \dot{\phi} + \partial_\phi V. \]  

(1.59)

We can define conditions on the evolution of the scalar field subject to its potential in terms of the slow-roll parameters

\[ \epsilon_v = \frac{1}{2} \left( \frac{\partial_\phi V(\phi)}{V(\phi)} \right)^2, \quad \eta_v = \frac{\partial_\phi \partial_\phi V(\phi)}{V(\phi)}, \]  

(1.60)

where we have adopted the lower index ‘V’ to track the definition with respect to the potential, rather than Hubble rate. Generally, slow-roll inflation occurs when \( \epsilon_v, |\eta_v| \ll 1 \); though the latter condition is not sufficient alone. These slow-roll parameters measure the shape of the potential at a given point in the fieldspace of \( \phi \), and in turn how supportive that may be for a slowly rolling scalar.

A convenient measure of how much the scale of the Universe grows throughout inflation is the number of exponential foldings, or \( e\)-folds, \( N \). This is defined as the ratio \( a_f/a_i = \exp(N) \), where the subscripts \( i \) and \( f \) denote evaluation at some initial and final time respectively. The integrated expansion is therefore

\[ N = \ln \frac{a_f}{a_i} = \int_{t_i}^{t_f} H dt. \]  

(1.61)

In the slow-roll limit, this can be computed in terms of the potential as

\[ N = -\int_{\phi_i}^{\phi_f} \frac{V}{\partial_\phi V} d\phi. \]  

(1.62)

Canonical values for the expansion such that classical big bang problems are resolved place \( 40 \lesssim N \lesssim 60 \). The precise value depends on the details of reheating, and how the Universe evolved thereafter.

In Fig. 1.4 we plot an example realisation of single field inflation with respect to a simple potential. When “inflating” we are in the slow-roll regime with \( \epsilon_H \ll 1 \) coloured in blue. Towards the end of the inflation \( \epsilon_H \to 1 \) as indicated in red, marking the end of inflation.

### 1.2.1.1 Perturbations in slow-roll

The evolution of a single scalar field is sufficient to describe a transient phase of exponential expansion. We now want to address how this leads to the formation of an initial curvature perturbation. The guiding principle to this generative process is that \( \phi \) is a quantum field. Locally, this will mean that \( \phi \) will fluctuate and be displaced from its background value, in the sense \( \phi(x, t) = \delta \phi(x, t) + \bar{\phi}(t) \). These fluctuations are small, such that we can expand
Figure 1.4: Realisation of single field inflation, in terms of its potential as a function of field value, $V(f_0) = m^2 f_0^2/2$. The inflationary trajectory starts at $f_0 \sim 17.5$ and ends near to $f_0 \sim 2$. The dynamical evolution is represented by the colouring of the line, starting blue for $\epsilon_H \ll 1$ and terminating when $\epsilon_H = 1$ in red. (Field units in $M_{Pl}$.)

the action Eq. (1.54) to second order and extract the equation of motion [47]

$$(\delta \phi)'' + 2H(\delta \phi)' + \nabla^2 \delta \phi + a^2 \partial_\phi \partial_\phi V \delta \phi = 0,$$  \hfill (1.63)

where for simplicity the FLRW background has been approximated as unperturbed; such that $\sqrt{-g} = a^4$.

Working in the slow-roll limit, we can drop the second derivative of the potential, so that in Fourier space

$$(\delta \phi_k)'' + 2H(\delta \phi_k)' + k^2 \delta \phi_k = 0,$$  \hfill (1.64)

Under canonical quantisation, the field perturbations are promoted to operators

$$\delta \hat{\phi}_k(\tau) = w(k, \tau) \hat{a}(k) + w^*(k, \tau) \hat{a}^\dagger(-k),$$  \hfill (1.65)

where $\hat{a}$ and $\hat{a}^\dagger$ are the creation and annihilation operators, the conjugate momentum is $\pi_{\delta \phi} = a^2 \delta \phi'$, and the functional coefficients $w, w^*$ provide solutions to Eq. (1.64). Their solution can be obtained analytically under the approximation that $H$ is slowly-varying enough that it is constant. If this is case, $\mathcal{H} = -1/\tau$ and

$$w'' - \frac{2}{\tau} w' + k^2 w = 0.$$  \hfill (1.66)

\footnote{For completeness, these are given by}

$$[\hat{\phi}(x, t), \hat{\pi}_\phi(x', t)] = i\delta(x - x'),$$  \hfill (1.67)

$$[\hat{\phi}(x, t), \hat{\phi}(x', t)] = 0,$$  \hfill (1.68)

$$[\hat{\pi}(x, t), \hat{\pi}(x', t)] = 0,$$  \hfill (1.69)

where $\pi_\phi = \partial L_\phi / \partial \dot{\phi}$ and $L$ is the Lagrangian of the action. The schematic calculation presented here loosely follows the ‘Wronskian’ method to obtaining the mode function.
Solutions for \( w \) are further restricted by the canonical commutation relations\(^7\) such that
\[
w(k, \tau) = \frac{H(k\tau - i)}{\sqrt{2k^3}} e^{-k\tau}.
\]
 Using this result we can compute the variance of the quantum fluctuations
\[
\langle 0 | \delta \hat{\phi}_k(\tau) \delta \hat{\phi}_{k'}(\tau) | 0 \rangle = w(k, \tau)w^*(k', \tau)\langle 0 | \hat{a}(k)\hat{a}^\dagger(-k') | 0 \rangle = |w(k, \tau)|^2 \delta(k + k'),
\]
with which we can identify the dimensionless power spectrum
\[
\mathcal{P}_{\delta \phi}(k, \eta) = \left( \frac{H}{2\pi} \right)^2 (k^2 \tau^2 + 1).
\]

We now want to relate this result to the comoving curvature perturbation, such that we can connect inflationary fluctuations to the initial conditions that evolve during the radiation era.

In the superhorizon limit \( k \ll H \leftrightarrow |k\tau| \ll 1 \) which suggests that
\[
\lim_{|k\tau| \to 0} \mathcal{P}_{\delta \phi} = \left( \frac{H}{2\pi} \right)^2.
\]

Now in order to relate this result to curvature, we recall that the comoving curvature perturbation can be defined in terms of energy density and pressure. In terms of the inflationary field, this can then be written as
\[
\mathcal{R} = -\frac{H}{\dot{\phi}} \delta \phi,
\]
indicating that we can associate a power spectrum of primordial curvature perturbations with the field fluctuations.

As previously advertised, the comoving curvature perturbation is conserved on superhorizon scales. This means that rather than taking the ‘full’ superhorizon limit as Eq. (1.73) might suggests, we should in fact evaluate the power spectrum at time of Horizon crossing. This leads to the classic single field result\(^8\)
\[
\mathcal{P}_R(k) \approx \left( \frac{H}{\dot{\phi}} \right)^2 \left( \frac{H}{2\pi} \right)^2 \bigg|_{aH=k}.
\]

This power spectrum encodes how the amplitude of scalar perturbations varies as a function of wavelength \( k \), which themselves identify with moments of Horizon exit in the inflationary expansion history. It is conventional to parametrise the scale dependence of

\(^8\)See e.g. Liddle & Lythe [48].
the power spectrum into the functional form

$$P_R(k) = A_s \left( \frac{k}{k_*} \right)^{n_s - 1},$$

(1.76)

where it is the spectral index, $n_s$, which actually absorbs the scale dependence. In general, the spectral index could contain various ‘runnings’, such that $n_s = n_s(k)$, and there are higher order derivatives of $n_s$ with respect to $k$. Usually this feature is ignored, since Eq. (1.76) turns out to provide an excellent fit over CMB scales; with the leading order runnings compatible with zero [49]. Despite being small, it may be possible in the near future to detect these runnings [50]. In order to compute $n_s$ we simply compute the logarithmic derivative of the power spectrum. In the slow-roll approximation, this can then be written in terms of the slow-roll parameters

$$n_s - 1 = -6\epsilon + 2\eta,$$

(1.77)

where evaluation at horizon crossing is implied.

![Figure 1.5: Primordial spectra (left) and their corresponding unlensed CMB angular spectrum (right) for a selection of representative spectral indices, $n_s$.](image)

As previously discussed, the scalar spectral parameters represent 2/6 of the base ΛCDM parameters. To get a handle on why these parameters are so important (and equally why they are well constrained), we simulate realisations of a primordial CMB surface for a representative choice of $n_s < 1$, $n_s = 1$ and $n_s > 1$ as per the spectra presented in Fig. 1.5. The corresponding last-scattering surfaces then follow in Fig. 1.6, in which we illustrate the relative size of fluctuations across scales.

![Figure 1.6: Simulated CMB surfaces with different primordial spectral indices. All simulations are full-sky and lensing-free, and are projected onto a 2D-Cartesian surface.](image)
1.2.2 Remarks on Gaussian random fields

Let us take amount to consider the Gaussian fields more generally. For a Gaussian field $G$, which is isotropic and homogeneous, the power spectrum is defined as

$$
\langle G(k_1)G(k_2) \rangle = (2\pi)^2 \delta(k_1 + k_2) P(k_1),
$$

(1.78)

where the pre-factors to $P(k)$ enforce isotropy. By definition, $P$ encodes the variance of a Gaussian field, and if that field is scale invariant, similar to what we advertised in the central frame of Fig. 1.6, then the power spectrum must satisfy $P(k) \propto k^{-3}$. This can be seen by demanding the statistical properties of $G$ to remain unchanged under a real-space linear transformation $x \rightarrow x' = \alpha x$, such that $\langle G(k'_1),G(k'_2) \rangle = \alpha^{-3} \langle G(k_1)G(k_2) \rangle$. As it turns out, the primordial power spectrum of the CMB, is almost scale-invariant; and slightly red. Primordial spectra of this type can be produced comfortably by inflation. The power spectrum can be made ‘dimensionless’ by rescaling the Gaussian power spectrum a cubic $k$ factor

$$
\langle G(k_1)G(k_2) \rangle = k^{-3} (2\pi)^2 \delta(k_1 + k_2) P(k_1),
$$

(1.79)

with which we define $P$ to be the dimensionless power spectrum.

For a purely Gaussian random field, it follows from Isserlis’ theorem [51] (a.k.a. Wick’s theorem) that all information is encoded in the 2-point correlation function. In particular, for higher order correlations

$$
\langle G(k_1),G(k_2),\ldots,G(k_p) \rangle = \begin{cases} 
\sum \prod_{i \neq j} \langle G(k_i),G(k_j) \rangle & \text{if } p = 2m \\
0 & \text{if } p = 2m + 1
\end{cases},
$$

(1.80)

for any integer integer $m$ greater than zero. This implies that the $n$-point correlator of a Gaussian random field can be described entirely by its power spectra. In single field inflation, the quantum fluctuations are Gaussian distributed and sourced by a single locally varying degree of freedom. Therefore as this Gaussianity maps into $\zeta$, we might expect no further information to be attained from studying higher order correlations.

For the study of other correlations to be worth while, we need $\zeta$ to somehow be non-Gaussian. The classic test-bed for assessing non-Gaussianity is the local model [52], parametrised by quadratic corrections to an otherwise Gaussian distributed curvature perturbation

$$
\zeta(x) = \zeta_G(x) + \frac{3}{5} f_{\text{local}}^\text{NL} (\zeta_G(x)^2 - \langle \zeta_G(x) \rangle^2).
$$

(1.81)

By construction, writing $\zeta$ in this way defines the amplitude of the non-Gaussian signal component to be $f_{\text{local}}^\text{NL}$. In Fourier space, the curvature perturbation equivalently can be written in as sum-separable Gaussian and non-linear corrective pieces $\zeta(k) = \zeta_G(k) + \zeta_{\text{NL}}(k)$.

Beyond the power spectrum, the first thing to consider is the Bispectrum, whose definition
is analogously given as
\[ \langle \zeta(k_1)\zeta(k_2)\zeta(k_3) \rangle = (2\pi)^3 \delta(k_1 + k_2 + k_3) B_\zeta(k_1, k_2, k_3). \] (1.82)

Direct computation with the local model for \( \zeta \) produces
\[ B_{\zeta}^{\text{local}}(k_1, k_2, k_3) = \frac{6}{5} f_{\text{NL}}^{\text{local}} (P_\zeta(k_1)P_\zeta(k_2) + P_\zeta(k_1)P_\zeta(k_3) + P_\zeta(k_2)P_\zeta(k_3)) \] (1.83)
where in the latter equality we have simply substituted the dimensionless power spectrum. Like the power spectrum, the Bispectrum can be made dimensionless via the momentum scaling
\[ B_\zeta(k_1, k_2, k_3) = B_\zeta(k_1, k_2, k_3)(k_1k_2k_3)^2. \] (1.84)

The dimensionless Bispectrum can be used to identify a shape, which along with its amplitude, turn out to be useful in forward modelling the expected CMB signal [53, 54, 55, 56].

In the ‘squeezed’ limit, the seminal result for single-field inflation is given by [57]
\[ \lim_{k_3 \to 0} \langle \zeta(k_1)\zeta(k_2)\zeta(k_3) \rangle = (2\pi)^3 \delta(k_1 + k_2 + k_3) P_\zeta(k_1)P_\zeta(k_2) \frac{d\ln P_\zeta(k_2)}{d\ln k_2}. \] (1.85)

Under identification of \( d\ln P_\zeta(k)/d\ln k = n_s - 1 \), it follows that for slow-roll models that \( n_s - 1 \sim \mathcal{O}(\epsilon) \ll 1 \). The bispectrum is vanishingly small, in a way characterised by slow-roll suppression. Geometrically, in the squeezed limit the smallest \( k \) acts as a background modulation to the shorter wavelengths. The correlation between the short scales appears as a power spectrum, and the long wavelength acts as a perturbation to the local background, which can be absorbed into a redefinition of the time (or scale factor).

### 1.2.3 The \( \delta N \) formalism

As a prelude to our following discussion, we will consider the evolution of super horizon perturbations. As per our discussion on conserved scalars, we argued that within an appropriately sized cosmological ‘patch’, a given perturbation could be absorbed into a definition of the local background. Ultimately this gave us the definition of the curvature perturbation as a means of understanding local density perturbations in the \( \delta \rho = 0 \) gauge.

Making use of this picture once again, we will consider patches of the Universe which are sufficiently large that spatial gradients become negligibly small, and we can consider each of these patches to evolve as independent FLRW universes. Since each patch is defined with its own set of initial conditions at some \( t_i \), at some future time \( t_f \) the local expansion will have differed from the background, and this difference can be interpreted as the uniform density perturbation. Denoting the background value, \( \bar{N} = \bar{N}(t) \), we will
denote the local expansion as

\[ N(x) = \int_{t_i}^{t_f} H(x, t) dt. \]  

(1.86)

From some initially flat hypersurface, we can then define

\[ \zeta = N(\phi^A) - \bar{N}, \]  

(1.87)

where \( \phi^A \) now represents the initial values for some number (\( A \)) of scalar fields. We will write each field as the sum of the background and perturbation value; \( \phi_A = \bar{\phi}_A + \psi_A \).

Assuming that the field perturbations are small, then any non-linear behaviour can be modelled accurately from a Taylor expansion

\[ \zeta = \sum_n \frac{1}{n!} N_{A_1 A_2 \cdots A_n} \psi^{A_1} \psi^{A_2} \cdots \psi^{A_n}, \]  

(1.88)

where for each term in the sum the Einstein summation is implied. The coefficients \( N_{A_1 A_2 \cdots A_n} \) are gradients of the local expansion with respect to the field fluctuations; e.g.

\[ N_{A_1} = \frac{\partial N}{\partial \psi^{A_1}}, \quad N_{A_1 A_2} = \frac{\partial^2 N}{\partial \psi^{A_1} \partial \psi^{A_2}}. \]  

(1.89)

If the fluctuations \( \psi \) are Gaussian random with zero mean, then they can be written as a series of the form

\[ \psi^A = \psi_1^A + \frac{1}{2} \psi_2^A + \frac{1}{6} \psi_3^A + \cdots, \]  

(1.90)

where the lower index serves as a form of bookkeeping, to track the order of in the expansion; with 1, 2, \( \cdots \) indicating the order of the field fluctuation. After expansion, it follows that only the leading order term is remain Gaussian random; which is simply Isserlis’ at work again.

We can identify the expected leading order behaviour of the \( n \)-point field correlations (a.k.a. the \( n \)-point functions, NPF) based on the rules of Gaussian statistics:

**2PF** Pure Gaussian pieces factorise into \( \langle \psi_1^A \psi_1^B \rangle \).

**3PF** Pure Gaussian pieces vanish, requires second order term \( \langle \psi_1^A \psi_1^B \psi_2^C \rangle \).

**4PF** Pure Gaussian pieces factories like 2pf, requires 2\(^{\text{nd}}\) and/or 3\(^{\text{rd}}\) order terms \( \langle \psi_1^A \psi_2^B \psi_3^C \psi_2^D \rangle \) and/or \( \langle \psi_1^A \psi_2^B \psi_3^C \psi_3^D \rangle \).

In the latter we are distinguishing between what is know as ‘disconnected’ (factorisable) and ‘connected’ contributions to the 4PF. Similar definitions hold for higher \( n \)-point functions.

Keeping track of the field fluctuations now appearing in terms of \( \zeta \), we have at first, second
and third order

\[
\begin{align*}
\zeta_1 &= N_A \psi_1^A, \\
\zeta_2 &= N_A \psi_2^A + N_{AB} \psi_1^A \psi_B^B, \\
\zeta_3 &= N_A \psi_3^A + N_{AB} (\psi_1^A \psi_2^B + \psi_2^A \psi_1^B) + N_{ABC} \psi_1^A \psi_1^B \psi_C^C.
\end{align*}
\]  

(1.91) (1.92) (1.93)

Working to leading order, the power spectrum is built from terms like \(\zeta_1\), the bispectrum from \(\zeta_1\) and \(\zeta_2\), and the trispectrum from \(\zeta_1\), \(\zeta_2\) and \(\zeta_3\). These quantities are formally defined respectively as

\[
\begin{align*}
\langle \zeta_{k_1} \zeta_{k_2} \rangle &= P(\zeta)(2\pi)^3 \delta(k_1 + k_2), \\
\langle \zeta_{k_1} \zeta_{k_2} \zeta_{k_3} \rangle &= B(\zeta)(2\pi)^3 \delta(k_1 + k_2 + k_3), \\
\langle \zeta_{k_1} \zeta_{k_2} \zeta_{k_3} \zeta_{k_4} \rangle &= T(\zeta)(2\pi)^3 \delta(k_1 + k_2 + k_3 + k_4),
\end{align*}
\]

(1.94) (1.95) (1.96)

where \(c\) denotes the connective trispectrum component, and each \(\zeta_{k_i}\) term appearing on the left hand sides correspond to a Fourier transformation of Eq. (1.88).

Retaining leading order correlations, it follows that the bispectrum can be written

\[
B(\zeta)(k_1, k_2, k_3) = \frac{6}{5} f_{NL}[P(\zeta)(k_1)P(\zeta)(k_2) + 2 \text{ perms.}],
\]

(1.97)

where

\[f_{NL} = \frac{5}{6} \frac{N_A N_B N_{AB}}{(N_C N_D)^2}.\]

(1.98)

and the trispectrum as

\[
T(\zeta)(k_1, k_2, k_3, k_4) = \tau_{NL}[P(\zeta)(k_1)P(\zeta)(k_2)P(\zeta)(k_3) + 11 \text{ perms.}] \\
+ g_{NL}[P(\zeta)(k_2)P(\zeta)(k_3)P(\zeta)(k_4) + 3 \text{ perms.}],
\]

(1.99)

where \(k_{ij} = |k_i + k_j|\) and

\[
\tau_{NL} = \frac{N_{AB} N^{AC} N^{B} N_C}{(N_D N^D)^3}, \quad g_{NL} = \frac{25 N_{ABC} N^{A} N^{B} N^{C}}{54 (N_D N^D)^3}.
\]

(1.100)

The bispectrum and trispectrum can be related via the inequality \([58, 59]\)

\[
\tau_{NL} \geq \left(\frac{6 f_{NL}}{5}\right)^2.
\]

(1.101)

In the single field limit, all perturbations are sourced by a single degree of freedom; and the non-linearity parameters simply depend upon the slow-roll parameters at horizon exit. In this case Eq. (1.101) has exact equality, \(f_{NL} \sim \mathcal{O}(\epsilon V)\) and \(\tau_{NL} \sim g_{NL} \sim \mathcal{O}(\epsilon^2 V)\), such that the non-linearities become vanishingly small.

Remarkably, with the \(\delta N\) formalism we can compute the expected non-linearities in \(\zeta\) from some initial field data. In practice, however, it can hard to implement. Most of this
difficulty stems from the many derivatives that need to be computed in order to assemble the \( \delta N \) expansion. The evolution of the expansion evolves necessarily on the details of the potential, and therefore the accessibility of this procedure depends on the models functional form.

In some instances, such as two-field models where the potential is sum-separable, \( \delta N \) can be used to obtain analytic solutions for observables [60, 61]. However, sum-separability is not a generic property of all potentials, and as such, it is difficult to define a universal procedure for obtaining analytic solutions more widely [62]. Notably, analytic computations typically rely on a number of approximations (such as slow-roll), which can be relaxed under a numerical implementation of the \( \delta N \) formalism [63]. This latter route offers a promising means of studying non-trivial models, and in particular, those described by multiple fields.

### 1.2.4 Multiple field inflation

**Extended notation.** We continue to use lower case Greek to denote all space-time metric components (e.g. \( \alpha \in \{0, 1, 2, 4\} \)) and lower case Latin for the transverse sector only (e.g. \( i \in \{1, 2, 3\} \)). We will now work with some number of fields \( N_f \) which we will denote with capital Latin counting from 1 (e.g. \( A \in \{1, \cdots , N_f\} \)). We will also work with phase-space vectors. We will use lower case sans serif to denote \( 2 \times N_f \) component phase-space vectors, spanning \( \{1, \cdots , N_f\} \) for the first component and \( \{N_f + 1, \cdots , 2N_f\} \) in the second (e.g. \( a \in \{1, 2, \cdots , 2N_f\} \)). Finally, Fourier integration is implied if contractions over field space indices are typeset in **bold sans serif**. For example,

\[
A_I B^I = \sum_I \int \frac{d^3k_I}{(2\pi)^3} A_I(k_I)B^I(k_I). \tag{1.102}
\]

As a starting point, let us extend the action in Eq. (1.54) to

\[
S = \int d^4x \left[ \frac{1}{2} R - \frac{1}{2} G_{AB}g^{\mu\nu} \partial_\mu \phi^A \partial_\nu \phi^B - V(\phi^A) \right]. \tag{1.103}
\]

In this action, we maintain an FRLW background cosmology, as defined by the the metric \( g_{\mu\nu} \). Instead of a single field, however, we now have a collection of \( N_f \) fields represented by \( \phi^A \), running over \( A = \{1, 2, \cdots , N_f\} \). These span what is referred to as a “field space”, which is controlled by the field space metric \( G_{AB} \). In the simplest case, the field space is Euclidean, and \( G_{AB} = \delta_{AB} \).

If the fields in Eq. (1.103) are homogeneous and time dependent, then the background equations of motion for the multi-field system are highly analogous to the single field case

\[
3H^2 = \frac{1}{2} \dot{\sigma}^2 + V, \quad D_t \dot{\phi}^A + 3H \dot{\phi}^A + G^{AB} \partial_B V = 0, \tag{1.104}
\]
where \( \frac{1}{2} \dot{\sigma}^2 = \frac{1}{2} G_{AB} \dot{\phi}^A \dot{\phi}^B = -\dot{H} \) is the kinetic energy of the fields. Here we have introduced the covariant time derivative, which acts on a field space vectors according to

\[
D_t \phi^A = \ddot{\phi}^A + \Gamma^A_{BC} \dot{\phi}^B \dot{\phi}^C,
\]  

where the Christoffel symbols should be assembled analogously to those in Eq. (1.9), but with the field space metric \( G \) in place of \( g \). In the case that the field space is flat, then \( D_t \dot{\phi}^A = \ddot{\phi}^A \), therefore the background evolution will intrinsically depend on the functional form of the field space metric. As in the single-field case, the criteria for inflation remains the same, as such, we define the slow roll parameter \( \epsilon_H = -\dot{H}/H^2 = \dot{\sigma}^2/2H \). To illustrate

**Figure 1.7:** Field space trajectory of a 2-field model of inflation, subject to the double quadratic potential

\[
V(f_1, f_2) = \frac{1}{2} m_1^2 f_1^2 + \frac{1}{2} m_2^2 f_2^2,
\]

for fields \( f_1 \) and \( f_2 \) with mass parameters \( m_1 \) and \( m_2 \). Starting inflation from some point high up the potential surface, the fields roll down towards the functional minima until \( \epsilon_H = 1 \), as drawn out by the red line. The background dynamics in this instance are computed with a Euclidean field space, \( G_{AB} = \delta_{AB} \). (Field units in \( M_{Pl} \).)

the background evolution in multiple field inflation, we extend the evolution presented in Fig. 1.4 to a 2-field scenario as given in Fig. 1.7. The potential is now represented by a surface, whose value is a function of the fields. We can immediately see from this simple example that inflation does not necessarily happen in a single “direction”, but has the ability to turn through the dimensions of field space. This turning can have implications for the perturbations sourced from field fluctuations, and is a distinguishing feature from the single-field scenario.
1.2.4.1 Covariant field fluctuations

As in the single field case, we will want to work with quantities of the form \( \phi^I \sim \bar{\phi}^I + \delta \phi^I \), where we decompose local field values in terms of their background and fluctuating components. The subtlety now is that the field space vectors \( \phi^I \) live on a field space manifold, \( G_{IJ} \), in which we wish to preserve the properties of tensor transformation. In defining the full field space vectors, this could present an issue, since they are manifestly coordinate dependent and therefore not covariant objects. For completeness, let us define the unperturbed field value at cosmic time \( t \), by \( \bar{\phi}^I(t) \), such that the fluctuating component of the field must then be the difference \( \delta \phi^I = \phi^I - \bar{\phi}^I \). This is a measure of a vector’s displacement in field space, that we want to construct in a covariant way.

To achieve this, we can connect the perturbed and homogeneous background vectors by a short and unique geodesic path, defined in terms of field-space metric \( G_{IJ} \). Under an affine parametrisation spanning some \( \lambda \in [0, \epsilon > 0] \), where evaluation at \( \lambda = 0 \) and \( \lambda = \epsilon \) correspond to the field space vectors \( \bar{\phi}^I \) and \( \phi^I \) respectively. These are essentially boundary conditions for the path connecting the unperturbed and perturbed vectors. To map between these points, we define the initial data

\[
\phi^I|_{\lambda=0} = \bar{\phi}^I, \quad \frac{d\phi^I}{d\lambda}|_{\lambda=0} = Q^I, \tag{1.106}
\]

where the vector \( Q^I \) lives in the same tangent space as \( \bar{\phi}^I \).

Expanding around \( \lambda = 0 \) and evaluating at \( \lambda = \epsilon \), the field space vector is expressible in terms of the series

\[
\phi^I(\lambda = \epsilon) = \frac{1}{n!} \frac{d^n \phi^I}{d\lambda^n} \bigg|_{\lambda=0} \epsilon^n. \tag{1.107}
\]

The geodesic equation for \( \lambda \) on \( G_{AB} \) is

\[
D^2_\lambda \phi^I = \frac{d^2 \phi^I}{d\lambda^2} + \Gamma^I_{JK} \frac{d\phi^J}{d\lambda} \frac{d\phi^K}{d\lambda} = 0. \tag{1.108}
\]

We can identify the \( n \)th order derivatives in this series expansion of \( \phi^I \) with the linear affine derivatives appearing in the geodesic equation and the boundary conditions of Eqs. (1.106).

The series expansion of the field space vector can then be written as

\[
\phi^I(\lambda = \epsilon) = \bar{\phi}^I + Q^I \epsilon - \frac{1}{2!} \Gamma^I_{JK} Q^J Q^K \epsilon^2 + \frac{1}{3!} (\Gamma^I_{LM} \Gamma^M_{JK} - \Gamma^I_{JLM}) Q^J Q^K Q^L \epsilon^3 + \cdots. \tag{1.109}
\]

With the freedom of the affine parametrization, let us now set \( \epsilon = 1 \), such that we can identify

\[
\delta \phi^I \equiv \phi^I - \bar{\phi}^I = Q^I - \frac{1}{2} \Gamma^I_{JK} Q^J Q^K + \frac{1}{3!} (\Gamma^I_{LM} \Gamma^M_{JK} - \Gamma^I_{JLM}) Q^J Q^K Q^L + \cdots. \tag{1.110}
\]

At linear order, we see that \( Q^I = \delta \phi^I \), and we have the canonical Euclidean expectation. This vector \( Q^I \) can be thought of as the covariant version of \( \delta \phi \), as such, we will just refer
to this as the field fluctuation (or perturbation) in the text that follows. For notational convenience, we will omit the over-bar used to define the background field space vectors, and simply use $\phi^I$ in a multi-field context.

In the previous discussion of single field inflation, we have adopted the slow-roll and ‘unperturbed metric’ approximations. This was a matter of convenience to derive the spectrum of curvature perturbations. In multiple field models this metric approximation is much less adequate. We will instead work within the ADM formalism \[65\] in order to keep track of how field fluctuations are communicated to the metric

$$ds^2 = -N^2 dt^2 + \gamma_{ij}(N^i dt + dx^i)(N^j dt + dx^j). \quad (1.111)$$

This can be considered analogous to the perturbed FLRW metric in Eq. (1.23), but now deviations from homogeneity have been absorbed into the lapse $N$ and shift $N^i$ functions. As before, $\gamma_{ij}$ represents the transverse components of the metric, and it assumes the form $\gamma_{ij} = a^2(t)\delta_{ij}$ on spatially flat hypersurfaces. This gauge choice is typical when working with multi-field analyses.

By writing the metric in terms of the extrinsic curvature\(^9\) $K_{ij}$, the action can be varied with respect to the lapse and shift functions in order to obtain a set of constraint equations \[66, 67\]. The constraints can be satisfied with first order expressions for lapse and shift, as given by \[68\]

$$N = 1 + \frac{1}{2H}G_{AB}D_t\phi^AQ^B,$$  \quad (1.112)

$$N_i = \frac{a^2}{2H}G_{AB}\left(Q^A D_t\phi^B - D_t\phi^A D_t Q^B - \frac{H}{H}D_t\phi^A Q^J\right).$$  \quad (1.113)

The constrain equations can then be fed back into the ADM metric.

To second order in $Q^I$, the action reads \[66\]

$$S_{(2)} = \int dt d^3x a^3 \left( G_{AB}D_tQ^A D_t Q^B - a^{-2}G_{AB}\partial_iQ^A\partial^iQ^B - M_{AB}Q^AQ^B \right),$$  \quad (1.114)

where $M_{AB}$ is known as the mass (squared) matrix, and couples the fields via

$$M_{AB} = V_{;AB} - R_{ACDB}\phi^C\phi^D - a^{-3}D_t\left(\frac{a^3}{H}\phi_B\phi_B\right),$$  \quad (1.115)

where $V_{;AB} = \partial_A\partial_BV - \Gamma_{AC}^{;B}\phi^C\partial_DV$ is the covariant Hessian matrix, and $R_{ABCD}$ is the Riemann tensor built from $G_{AB}$. The equations of motion for the fluctuations are then given by \[69\]

$$D_t D_t Q^A + 3HD_t Q^A + \left[\frac{k^2}{a^2}\delta_B^A + M_B^A\right] Q^B = 0.$$

\(^9\)Explicitly, this is given by the combination $K_{ij} = \frac{1}{2N}(\partial\gamma_{ij} - N_{ij;j} - N_{ji;i})$, where ‘;’ refers to the covariant derivative with respect to $\gamma_{ij}$. Under an analogous procedure, it is possible to further obtain the action to third order in
perturbations.

In order to relate the field fluctuations to the curvature perturbation, we do so via considering the energy density [70]. Assuming that it is only the inflationary fields that populate the Universe, then the EM tensor can be computed directly from the field space action, and therefore will depend on the field content $T_{\mu\nu} = T_{\mu\nu}(G_{AB}, \phi^A)$. In general, the energy density is obtained from the temporal sector of the EM tensor via $\rho = -T^{00}/g^{00}$, such that at the background level $\rho = \frac{1}{2}G_{IJ}\sigma^2 + V(\phi^I)$, which is simply the multi-field generalisation to the single field result. More interestingly, the perturbed energy density can be written in terms of the perturbed fields. This enables the perturbed energy density to be directly related to the curvature perturbation of constant density. This can be done to arbitrary order to obtain, e.g. $\zeta^{(1)} \sim Q^I$, $\zeta^{(2)} \sim Q^I Q^J$, $\cdots$ [70, 71].

We want to quantise these fluctuations, in order to compute the spectrum of curvature perturbations. However, approaching this problem is more subtle for multiple fields. The primary concern is the evolution over superhorizon scales. In the single field paradigm, we have only one locally varying degree of freedom, which is known to be conserved on superhorizon scales. With multiple fields this is no longer true, and therefore the power spectrum could change between horizon exit and re-entry [72]. Under certain conditions, this exchange between ‘adiabatic’ and ‘entropic’ modes can be controlled, but typically relies on assumptions about the states of the fields on horizon exit [73]. The situation can become complicated if the fields fluctuate differently across, sub-Hubble, Hubble and super-Hubble scales, such that a form of analytic-continuation or a numerical formalism is required [74].

Most analytic methods operate on super-horizon scales, adopting a ‘separate universe’ approach to understanding the dynamical evolution. The classic formalism is $\delta N$ as discussed in Sec. 1.2.3, which makes use of the gradients of the inflationary fields measured against the expansion history, in order to compute the late time correlations of $\zeta$ using data on the early time correlators of $\delta \phi$. In the following section we turn our attention to a different formalism that focusses on the numerical solution for inflationary correlation functions. This formalism is exercised at length in the publication ‘Non-Gaussianity in D3-brane inflation’ appearing as Chpt. 2.

1.2.5 The ‘Transport’ approach

In order to compute the two point correlation function we will need to use the second order action for the perturbations. Additionally, we will consider the three-point correlator, such that we will also need to the third order action. Making use of the notation defined in
Eq. (1.102), these actions can be written in Fourier space as

\[
S^{(2)} = \frac{1}{2} \int dt a^3 \left[ G_{IJ}(\mathbf{k}_I, \mathbf{k}_J)(D_t Q^I(\mathbf{k}_I))(D_t Q^J(\mathbf{k}_J)) 
\right.
\]
\[+ M_{IJ}(\mathbf{k}_I, \mathbf{k}_J)Q^I(\mathbf{k}_I)Q^J(\mathbf{k}_J),
\]

\[
S^{(3)} = \frac{1}{2} \int dt a^3 \left[ A_{IJK}(\mathbf{k}_I, \mathbf{k}_J, \mathbf{k}_K)Q^I(\mathbf{k}_I)Q^J(\mathbf{k}_J)Q^K(\mathbf{k}_K)
\right.
\]
\[+ B_{IJK}(\mathbf{k}_I, \mathbf{k}_J, \mathbf{k}_K)D_t \{Q^I(\mathbf{k}_I)\}Q^J(\mathbf{k}_J)Q^K(\mathbf{k}_K)
\]
\[+ C_{IJK}(\mathbf{k}_I, \mathbf{k}_J, \mathbf{k}_K)D_t \{Q^I(\mathbf{k}_I)\}D_t \{Q^J(\mathbf{k}_J)\}Q^K(\mathbf{k}_K)\right].
\]  

The actions have acquired Fourier ‘coefficient tensors’ when moving from their real-space representation. These are all proportional to delta-functions, and some combination of data obtainable from the field-space background evolution. For instance, in the second order action

\[
G_{IJ} \propto G_{IJ},
\]

and

\[
M_{IJ} = (2\pi)^3 \delta(\mathbf{k}_I + \mathbf{k}_J) \left( \frac{\mathbf{k}_I \cdot \mathbf{k}_J}{a^2} G_{IJ} - M_{IJ} \right),
\]

which can be written in terms of mass-matrix from Eq. (1.115). The remaining coefficients are much more algebraically complex, so have been omitted for brevity. These quantities define what is referred to in the transport literature\(^{10}\) as the \(M\)-, \(A\)-, \(B\)- and \(C\)-tensors.

By variation of the action with respect to the covariant time evolution of the perturbations, we can obtain their canonical momentum

\[
P_A = \frac{\delta S}{\delta(D_t Q^A)},
\]

which in term can also be expressed in terms of the coefficient tensors. It is useful to collect the perturbations and their momenta into a phase-space vector, which we will denote as \(\delta X^a = (Q^A, P^B)\), where \(a \in \{0, 1, \ldots, 2 \times N_t - 1, 2 \times N_t\}\).

From the actions Eqs. (1.117, 1.118), we can obtain the equations of motion for the perturbations. On subhorizon scales, the perturbations will behave quantum mechanically, and therefore may be interpreted as Heisenberg operators. These operators satisfy Hamilton’s equations, such that \(^{11}\)

\[
D_t Q^I = -i[Q^I, \mathcal{H}],
\]
\[
D_t P^I = -i[P^I, \mathcal{H}] - 3H P^I,
\]

where \(\mathcal{H}\) is the Hamiltonian density\(^{11}\), which is generally a combination of free and interacting parts.

\(^{10}\)For Euclidean field space these expressions can be found in Ref. [75], and their extension to curved field space in Appendix A of Ref. [76] or Ref. [77].

\(^{11}\) The explicit Hamiltonian can be computed via integration over the ‘Hamiltonian density’ of the system. This is analogous to the standard definition for canonical scalar fields, but now the system is specialised to the inflationary phase-space. Direct computation shows that the momentum can be written in terms of the coefficient tensors. See Eq. 2.4.5 of Ref. [77] for an explicit expression.
Taking the equations of motion for the Heisenberg operators, the expectation values of products over such operators can be obtained via Ehrenfest’s theorem [75, 78].

From the combined phase space vector of perturbations, the associated 2- and 3-point correlation functions are given by

\begin{align}
\langle \delta X^a(k_a) \delta X^b(X_b) \rangle &= (2\pi)^3 \delta(k_a + k_b) \Sigma^{ab}(k_a), \\
\langle \delta X^a(k_a) \delta X^b(X_b) \delta X^c(X_c) \rangle &= (2\pi)^3 \delta(k_a + k_b + k_c) B^{abc}(k_a, k_b, k_c). 
\end{align}

From Ehrenfest’s theorem and Hamilton’s equations, the time evolution of the expectation values is governed by

\begin{align}
D_t \Sigma^{ab}(k) &= u^a_c(K) \Sigma^{cb}(k) + u^b_c(K) \Sigma^{ac}(k), \\
D_t B^{abc}(k_a, k_b, k_c) &= u^a_d(k_a) B^{dbc}(k_a, k_b, k_c) \\
&\quad + u^b_d(k_b) B^{adc}(k_a, k_b, k_c) \\
&\quad + u^c_d(k_c) B^{abd}(k_a, k_b, k_c) \\
&\quad + u^a_{de}(k_a, -k_b, -k_c) \Sigma^{db}(k_b) \Sigma^{ec}(k_c) \\
&\quad + u^b_{de}(k_b, -k_a, -k_c) \Sigma^{ad}(k_a) \Sigma^{ec}(k_c) \\
&\quad + u^c_{de}(k_c, -k_a, -k_b) \Sigma^{ad}(k_a) \Sigma^{be}(k_c).
\end{align}

The covariant time derivatives acting on the operators induce an extended form of connexion components, with the rules

\begin{align}
D_t \Sigma^{ab}(k) &= \partial_t \Sigma^{ab}(k) + \Gamma^a_c(k) \Sigma^{cb}(k) + \Gamma^b_c(k) \Sigma^{ac}(k), \\
D_t B^{abc}(k_a, k_b, k_c) &= \partial_t B^{abc}(k_a, k_b, k_c) + \\
&\quad + \Gamma^a_d(k) B^{dbc}(k_a, k_b, k_c) \\
&\quad + \Gamma^b_d(k) B^{adc}(k_a, k_b, k_c) \\
&\quad + \Gamma^c_d(k) B^{adb}(k_a, k_b, k_c),
\end{align}

where

\[ \Gamma^g_b = \begin{pmatrix} \Gamma^f_{JK} \phi^f & 0 \\ 0 & \Gamma^f_{JK} \phi^{\dagger f} \end{pmatrix}. \]
The associated $u$-tensors are given as
\[ u^a_b = \begin{pmatrix} 0 & \delta^I_J \\ \hat{M}^I_J & -3H\delta^I_J \end{pmatrix}, \] (1.130)
\[ u^a_{bc} = \begin{pmatrix} -b^I_{JK} & -c^I_{JK} \\ 3a^I_{JK} & b^I_{KJ} \\ -c^I_{KJ} & 0 \\ b^I_{JK} & c^I_{KJ} \end{pmatrix}, \] (1.131)
where $\hat{M}_{IJ} = -k^2G_{IJ}/a^2 - M_{IJ}$.

Note that in general the two-point function is complex, in the sense that
\[ \Sigma^{ab} = \Sigma^{ab}_{\text{Re}} + i\Sigma^{ab}_{\text{Im}}. \] (1.132)

This means that the equations of motion for the real and imaginary parts of the 2-point satisfy Eq. (1.125) independently, and moreover, the imaginary component decays on super horizon scales. This is essential for the large-scale inflationary perturbations to satisfy classical equations of motion. The 3-point correlator is in general also imaginary, but is manifestly real when restricted to tree-level (which is the case for the current Transport implementation). After finding the real and complex solutions for the 2-point correlator in Eq. (1.132), terms in the 3-point correlator that are quadratic in the two-point correlation are expanded as
\[ \Sigma^{db}\Sigma^{ec} \rightarrow \Sigma^{db}_{\text{Re}}\Sigma^{ec}_{\text{Re}} - \Sigma^{db}_{\text{Im}}\Sigma^{ec}_{\text{Im}}, \] (1.133)
i.e. retaining only real terms.

### 1.2.5.1 Initial conditions

In order to specify initial data to evolve the correlation functions forwards from, we turn to the subhorizon regime, long before scales of interest have exited the horizon.

With respect to the 2-point function, at early enough times, their description will approach that of a system of massless uncoupled scalar fields [75]. To identify a sufficiently early time when this 'massless condition' is attained, we can compare the largest eigenvalue $m^2$ associated with the mass spectrum $M^I_J$ to a Fourier mode of interest\(^{13}\) $k$

\[ m^2 \ll \frac{k^2}{a^2}. \] (1.134)

This analytic criteria arises by comparing terms in the Fourier coefficient tensor $M_{IJ}$ appearing in the second order action Eq. (1.119). This inequality should\(^ {14}\) be satisfied at times early enough in the expansion history, long before $k$ has exited the horizon. In

\(^{13}\) An ‘interesting’ mode $k$ is an appropriate pivot scale which should be of observational relevance, e.g. to characterise CMB anisotropies.

\(^{14}\) There are instances where this condition may not be attained, which correspond to realisations of
practice, we devise a criteria based on a fixed number of e-folds before being massless

\[ N_\ast \equiv N_{\text{Massless}} - N_\delta : \quad m^2 \bigg|_{N_{\text{Massless}}} = \frac{k^2}{a^2} \bigg|_{N_{\text{Massless}}}, \quad (1.135) \]

for \( N_\delta \geq 0 \). The larger \( N_\delta \) the more accurate the initial conditions will become, at the expense of tracking a longer duration of sub-horizon evolution; which can become computationally prohibitive.

Assuming we have found a sufficiently early time, we wish to now compute the corresponding 2-point correlation of perturbations \( Q^I \). This calculation is analytically challenging, but is greatly simplified, by utilising the approximate (though highly accurate) de Sitter spacetime solution [66].

In the limit of equal time correlation, the initial conditions for the initial \( N_\ast \) can be written Ref. [79, 77]

\[
\Sigma_{ab} \text{ Re} = \frac{1}{2a^3k} \begin{pmatrix}
 aG^{IJ} & -aHG^{IJ} \\
 -aHG^{IJ} & (k^2/a)G^{IJ}
\end{pmatrix},
\]

\[
\Sigma_{ab} \text{ Im} = \frac{1}{2a^3k} \begin{pmatrix}
 0 & aG^{IJ} \\
 -kG^{IJ} & 0
\end{pmatrix}.
\]

Constructing the initial conditions for the 3-point data is more involved, with large non-trivial expressions. We refer the reader to Ref. [75] for the canonical field space calculation, and in particular, Section 4 and Appendix A.1 of Ref. [77] regarding the more general curved field space computation.

1.2.5.2 Connection to \( \zeta \)

We need to compute the correlations of \( \zeta \) to make observational predictions. Therefore we need to relate \( \zeta \) to the field perturbations \( \delta X^a \).

On super-horizon scales, this relationship takes the form

\[ \zeta(k) = N_a \delta X^a + \frac{1}{2} N_{ab} \delta X^a \delta X^b, \quad (1.138) \]

where

\[ N_a(k) = (2\pi)^3 \delta(k - k_a) N_a, \quad (1.139) \]

\[ N_{ab}(k, k_\alpha, k_\beta) = (2\pi)^3 \delta(k - k_\alpha - k_\beta) N_{ab}(k_\alpha, k_\beta). \quad (1.140) \]
The canonical 2- and 3-point correlation functions of $\zeta$ are defined by

$$\langle \zeta(k_1)\zeta(k_2) \rangle = (2\pi)^2 \delta(k_1 + k_2) P(k),$$

(1.141)

$$\langle \zeta(k_1)\zeta(k_2)\zeta(k_3) \rangle = (2\pi)^2 \delta(k_1 + k_2 + k_3) B(k_1, k_2, k_3),$$

(1.142)

where the power spectrum and bispectrum in the Transport formalism now take the form

$$P(k) = N_a N_b \Sigma_{Re}^{ab}(k)$$

(1.143)

$$B(k_1, k_2, k_3) = N_a N_b N_c B^{abc}(k_1, k_2, k_3) + \left[N_a N_b N_{cd}(k_1, k_2) \Sigma_{Re}^{ac}(k_1) \Sigma_{Re}^{bd}(k_2) + 2 \text{ cyclic}\right].$$

(1.144)

As shown in Ref. [70], the superhorizon curvature perturbation can be related to the total density perturbation on flat hypersurfaces by

$$\zeta = -H \frac{\delta \rho}{\rho} + H \frac{\delta \rho}{\rho^2} - H^2 \frac{\delta \rho^2}{2 \rho^3} + \dot{H} \frac{\delta \rho^2}{2 \rho^2}.$$

(1.145)

In general, computing $\rho$ depends on the specification of $T_{\mu\nu}$ in the cosmological model, and therefore for the case at hand is derived from a universe filled with scalar fields minimally coupled to gravity. Variation in the stress-energy can be expressed in terms of the field fluctuations $\delta \phi$ which in turn can be traded for their covariant counter part $\delta Q$. This enables the curvature perturbation to be expressed in terms of linear and quadratic pieces, e.g. [77]

$$\zeta^{(1)} = -\frac{1}{2 M_P^2 H \epsilon} G_{IJ} \dot{\phi}^I Q^J,$$

(1.146)

whilst the second order term depends on quadratic pairs of $Q^I Q^J$, as well as products of covariant time derivatives $Q^I \dot{D}_J Q^J$ (see Eq. 5.2.2 of Ref. [77]).

Under Fourier transformation, expressions for the $N-$tensors are found in terms of the homogeneous background, e.g.

$$N_a = -\frac{1}{2 M_P^2 H \epsilon} \dot{\phi}^I \begin{pmatrix} 1 \\ 0 \end{pmatrix},$$

(1.147)

with $N_{ab}$ similarly relying only on background level data (see Eq. 5.2.2 of Ref. [77]).

### 1.3 Sampling with PyTransport

PyTransport [80] is a software implementation of the aforementioned ‘Transport’ approach to inflationary correlation functions. For a given inflationary model, the software ingests symbolic Python expressions\(^{15}\) for the potential (and field space metric, if relevant), customises a suite of C++ template sources defining the integral solutions to inflationary

---

\(^{15}\)Written in SymPy [81].
correlation functions, before compiling them into an extended Python module. In this way, PyTransport capitalises upon the numerical efficiency of C++, whilst retaining the highly flexible Python interface.

Whilst the PyTransport technology is readily deployable for model analysis with respect to a discrete choice of parameters, a robust exploration of observables with respect to a given model invites an interface with sampling technology. This is advocated as to avoid tedious (and likely erroneous) manual redefinitions, and to ensure coherency across computations with respect to analysis procedures and cosmological assumptions.

Such developments are packaged within the PyTransport Sampler\(^{16}\) (pytsa) repository. A version of this technology was deployed in Chpt. 2 to study the D3-brane inflationary scenario, whilst further developments have been incorporated into the latest release.\(^{17}\)

The software can execute the inflationary analyses with two parameter scripts: one script to install the model, and another to assemble the sampling framework. Example scripts highlighting the core implementation are deferred to Appendix A.

### 1.4 Context for original research

#### 1.4.1 Probing high-energy theory

As we understand it, inflation took place somewhere between energies of \(1 \text{ MeV} \lesssim E_{\text{inflation}} \lesssim 10^{15} \text{ GeV}\). This enormous prospective range of energies is conservative, though characterises well motivated physical principles. The lower bound corresponds to those associated with nucleosynthetic processes (i.e., the binding energy of Hydrogen nuclei), whilst the upper by the expected scales of Grand Unification Theory (GUT): where the weak, strong, and electromagnetic forces are thought to become unified.\(^{18}\)

Beyond these energy scales things become additionally speculative. However, it is largely assumed that there was a Planck epoch where the effects of quantum gravity are important, and before which, some initial singularity. Classical bids to extend general relativity towards Planck scale physics “break down”, in the sense that when interpreted as a field theory, coupling strengths are characterised by Newtons constant \(G\), which grow in strength as the energy scale is increased \(\cite{83}\). This effect is non-renormalisable, and leads to an intrinsic divergence.\(^{19}\) This motivates the postulation that additional physics must

---

\(^{16}\)https://github.com/bkmarzouk/pytsa

\(^{17}\)Most of these developments pertain to the stability of software itself: full alignment with Python 3 and deprecation of Python 2, reproducibility of sampling analyses, write-safe procedures to mitigate accidental sample contamination, more robust definitions for prior sampling range, unit testing, and package restructuring for more ‘Pythonic’ extended module importation. With respect to the latter update, previous versions of PyTransport required manual path configuration in order to import extended C++ modules, whereas in this latest release this is straightforwardly achieved via the standard syntax: from pytsa.models import model

\(^{18}\)See Georgi & Glashow for the classic paper on this topic \(\cite{82}\).

\(^{19}\)In gravity only theories, a precise cancellation at the 1-loop level prevents divergence. Unfortunately,
exist towards the Planck scale, as to reconcile gravity at high-energy scales. A popular interpretation is that the ultraviolet (UV) theory describing this regime should produce general relativity in the low-energy limit; or at least something that resembles it close enough to pass the observational tests. This interpretation is that General Relativity is therefore an effective field theory [85].

The conventional approach to probing high-energy physics is through particle colliders, such as the Large Hadron Collider (LHC). Searching for new physics with colliders is experimentally challenging, though ideally straightforward: By accelerating particles to high enough energies there is a chance that their net energy can stimulate otherwise unobservable interactions. Such interactions may produce exotic particle content, or decay configurations that are only explainable with ‘new’ physics. These methods recently vindicated the Higgs Bosons place within the standard model [86, 87]. For decades, it has been further hoped that there would be evidence of ‘Super Symmetry’ [88] - a popular extension to the ‘Standard Model’ - though in the absence of evidence, proponents of the extension have assumed that it must only be relevant at every-increasing energy scales.

It is unlikely that quantum gravitational effects will be observable with terrestrial collider technology. Confoundingly, one of the best prospects of identifying observables for UV physics is through its relationship with another speculative theory: cosmic inflation [69].

To entertain this possibility, we must assume that the UV complete theory is connected to inflation in some form of low-energy limit. In doing so, there are then two ways in which we could try to understand the relationship. From the ‘bottom-up’: we could consider the (relatively) low-energy theory of inflation, and work with a set of guided assumptions on the energy scales at which symmetries and operators from the UV sector become relevant. From the ‘top-down’: we could start with the UV theory, and derive expected behaviours in the low-energy limit. When working in either of these two directions we are trying to formulate an ‘effective field theory’ description.

From a theoretical mindset, let us propose that within a certain low-energy limit of some UV complete theory that a viable model of inflation can and does arise. It is conceivable that constraints on the late-time cosmological signatures associated with the child inflationary model would translate into constraints on the parent UV theory. Prospectively, things could be even better, if it turns out that the cosmological signature is so distinctly recognisable (e.g. a large-amplitude and scale-dependent non-Gaussianity) that it would be very hard to explain by any other means of inflation. Ex post facto, one could claim supporting evidence of UV physics from cosmological observations.

For string-theory aficionados, this presents the exciting prospect of obtaining ‘experimental’ data to confront the theory space. Indeed, if string-theory is the correct theory of quantum gravity, it should be able to explain the emergence of inflation in a predictable way.

---

this is not generally true for theories including matter. [84]
A key difficulty in realising inflation from string theory is obtaining an appropriate form of space-time that is able to describe an inflationary cosmology. As such, important areas of research have targeted the pursuit of de Sitter vacua from string theory. Famously, a viable mechanism for generating such vacua was shown by ‘KKLMMT’ [89], with the assistance of extended topological objects known as (anti) ‘D3-branes’ as stabilising components. With the addition of a mobile D3-brane, this de Sitter vacua solution can support a highly computable variant of d-brane inflation (a proposition we discuss at length in Chpt. 2). This scenario is schematically represented in Fig. 1.8 (see Fig. 2.1 later for a more complete picture20). The key driver for inflation in this scenario is the oppositely charged brane/anti-brane pair, which exhibit a Coulomb-like attraction. This attraction induces a fieldspace excursion through fieldspace with respect to the mobile brane. This simple picture is complicated by non-trivial interactions with the “bulk”, which can propel or pull the brane back and forth. At a glance, this additional complexity may seem like a curse, however, it is through coincidental balancing that a suitable scalar potential can be constructed to support a viable period of inflation; which would otherwise be too steep based on the Coulombic interaction alone.

1.4.2 Constraining primordial non-Gaussianity

As we have previously outline: the primordial CMB structure laid down by a curvature power spectrum generated by inflation seems to be well described as a Gaussian random field, characterised by deviances from a mean background radiation temperature $\delta T/\bar{T}$. This is similarly true in polarisation.

We have also emphasised that inflation can produce non-Gaussian correlations in $\zeta$ as represented by a non-zero amplitude in the bi- and trispectra, which can occur if there are multiple fields driving the expansion.21

---

20The text about this figure systematically describes all of the relevant details of the brane model which have been omitted here for brevity.

21So where could these multiple fields come from? The classic proposition is that scalar fields are ubiquitous to string-theory, in which multiple field inflation is expected to be generic. The aforementioned d-brane scenario can in fact be interpreted as a system of 6 scalar fields, which parametrise the coordinate separation of the branes. This is explained in detail in Chpt. 2.
We can understand how this connects to our observations of the CMB from Eq. (1.48), where we wrote the temperature harmonics in terms of some integral over the primordial curvature perturbation and transfer function: \( a_{\ell m} \sim \int T(\ell)(\log(1 + \ell k)) \sigma_{\ell m}(k) dk/k \). Therefore, by analysing the \( n \)-point correlation functions of harmonic data, we might be able to learn about the \( n \)-point primordial correlations.

To this end, there has been significant effort invested into understanding the CMB poly spectra, and in particular, the bispectrum. (This is primarily due to the expectation that it contains more signal to noise than the trispectrum or higher.) In a statistically isotropic way, the CMB bispectrum can be written generally as

\[
\langle a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3} \rangle = B_{\ell_1 \ell_2 \ell_3} \left( \frac{\ell_1}{m_1} \frac{\ell_2}{m_2} \frac{\ell_3}{m_3} \right) = b_{\ell_1 \ell_2 \ell_3} \int d\Omega Y_{\ell_1 m_1} Y_{\ell_2 m_2} Y_{\ell_3 m_3}, \tag{1.148}
\]

where \( B_{\ell_1 \ell_2 \ell_3} \) and \( b_{\ell_1 \ell_2 \ell_3} \) are the CMB bispectrum and reduced-bispectrum respectively.

In order to estimate \( f_{\text{NL}} \) from the CMB bispectrum, the general approach is to forecast the expected signal based on some fiducial bispectrum shape, and attempt to fit the CMB bispectrum observation to that template. There are many estimators in the literature, see for example Refs. [90, 91, 92, 94, 95, 96, 97]. Trispectrum estimators can be formulated in a similar way [98, 99, 100], though they prove quite difficult to use in practice (partly due to combinatorics).

The tightest constraints on the bispectrum are from the Planck collaboration, who are found \( f_{\text{NL}} \) to be consistent with zero in a range of \( k \) configurations, as shown in Tab. 1.2 [27]. Generally, the local signal is the easiest to constrain on CMB scales and is therefore the most statistically significant using current state-of-the-art cosmological data [101]. Notably, other shapes may be more easily measured using 3D survey data [102].

| \( f_{\text{NL}}^{\text{local}} \) | -0.9 ± 5.1 |
| \( f_{\text{NL}}^{\text{equil}} \) | -26 ± 47 |
| \( f_{\text{NL}}^{\text{ortho}} \) | -38 ± 24 |

Table 1.2: Constraints on the reduced-bispectrum from Planck 2018. Top to bottom represents constraints at the 68% C.L. for the (squeezed) local, equilateral and orthogonal configurations.

Notwithstanding the bispectra amplitudes’ consistency with zero, there remains viable windows for prospective future detection, with a clear motivation for experimental configurations to target sensitivities of \( \sigma(f_{\text{NL}}) \sim \mathcal{O}(1) \).

Beyond the bispectrum, we can constrain the trispectrum. As illustrated by the \( \delta N \) formalism, the amplitude of \( \tau_{\text{NL}} \) is related via inequality to \( f_{\text{NL}} \). This means that there is a possibility of measuring a significant trispectrum whilst in the absence of a large bispectrum.

As mentioned with respect to template estimation, estimating the trispectrum is generally
difficult. This is primarily due to the low-signal to noise, coupled with the large number of mode configurations that are required to apply direct estimators. An alternative approach for the assessment of this squeezed local non-Gaussianity is founded in the ‘modulation’ formalism. In this formalism, we model the largest-scale curvature perturbation associated with a squeezed local trispectrum as approximately constant through the shell of last-scattering. These large scale perturbations induces spatial variations in the otherwise statistically isotropic surface at last scattering via their modulation of small scale perturbations. This parametrises the problem in a more tractable way, in which we can look for signatures of statistical anisotropy. Methods for estimation of this type are mature, and have been successfully applied to estimate the CMB lensing field.

In terms of a primordial trispectrum, we are looking for what is referred to as a ‘modulation field’. In Fig. 1.9 we illustrate what the modulated signal component might look like for a squeezed $\tau_{NL}$-like signal, which we discuss in detail in Chpt. 3.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1.9.png}
\caption{Real space representation of the modulated component of a CMB temperature field, $T(\hat{n})$, subject to a squeezed local trispectrum non-Gaussianity. Neglecting noise, the full CMB temperature signal is modelled as $T(\hat{n}) = (1 + f(\hat{n}))T_g(\hat{n})$, where $f(\hat{n})$ is the modulating field and the subscript $g$ denotes Gaussian. Position space variations in power can clearly be seen by eye.}
\end{figure}
Chapter 2

Non-Gaussianity in D3-brane inflation

Kareem Marzouk, Alessandro Maraio and David Seery  \textit{JCAP02(2022)013}

Inflation continues to retain its favoured position as the leading scenario for the origin of structure in the universe—but there has been little progress towards identifying the degrees of freedom that were active during the inflationary era, or the manner in which they interacted. Among the major reasons for this slow progress are well-rehearsed arguments showing that inflation is sensitive to small, nonrenormalizable interactions suppressed by Planck-scale masses, and therefore may depend on the precise way in which the inflationary sector is embedded within its ultraviolet completion.

This phenomenon is not a failure of decoupling in the technical sense \cite{103}, but shares its double-edged character.\footnote{Let us assume a high-energy theory which is known to be valid down to some low-energy cut off. We could build a low-energy version of such a theory, for example, by ‘integrating out’ the high-energy interactions, i.e. those occurring above the low-energy cut off scale, and expanding the resulting low-energy action in terms of a series of operators and coupling constants. By repeating this process, one can create a cascading stack of theories which correspond to lower and lower energies. Theories that are valid at higher energies become effectively ‘decoupled’ from those below in this way.} On the one hand, if successful inflation can depend on physics at or near the Planck scale, we are encouraged to believe that it may be possible to discover details of quantum gravity by studying inflationary observables. On the other, dependence on high-scale physics means that inflation can \textit{not} be studied on its own: assumptions about physics at higher energies are required, even if they are not made explicit. The predictivity of the scenario is therefore reduced.

This situation has encouraged development of approaches in which inflationary model-building takes place in the context of a concrete proposal for its ultraviolet completion. The most well-developed of these use string theory as the ultraviolet model, paired with a variety of suggestions for the microscopic origin of the low-energy fields that populate the inflationary sector. The field was surveyed at length in a recent monograph by Baumann & McAllister \cite{69}.

One proposal is that inflation is driven by the dynamics of a D3$/\overline{D3}$ brane pair within a warped deformed conifold. The attraction of this scenario is not that we think it more realistic than any other model, but that it is highly computable. In particular, the functional form of the low-energy effective action can be computed reliably, even accounting for
contributions from moduli stabilization and supersymmetry breaking. This high degree of computability is remarkable.

The most significant drawback is the complexity of the resulting effective theory: the potential we describe in §2.1 below has 1,212 independent parameters and is a sum of 3,881 terms, many of which are themselves complicated. Numerical methods are necessary: it is impractical to extract observational predictions from such complex potentials using analytical techniques. Progress therefore becomes dependent on compute resource and the availability of suitable software tools. It is arguable that analysis of this model—and others of comparable complexity—has been hampered by both the paucity of powerful, general-purpose software tools for inflationary model analysis, and an accepted means for exchanging the specification of models within the community. In the sister discipline of collider phenomenology these roles are played by the FeynRules system and its online model database [104, 105].

These disadvantages notwithstanding, its unusual theoretical control has made the D3/D3 model an interesting laboratory in which to study the likelihood of inflation, the distribution of observables such as the primordial spectral index, and the prospects for accommodating fine-tuning issues such as the well-known ‘η-problem’ (that is, light scalar fields in a quasi-de Sitter spacetime typically acquire masses of order $H$). For these reasons the model has developed its own literature, which we review briefly beginning on p. 45.

Non-Gaussianity and observables.—In this paper we return to the D3/D3 model and reanalyse it using updated numerical methods. Our principal aim is an accurate characterization of the primordial non-Gaussianity it produces, for which reliable estimates have not yet been reported. To achieve this we leverage new versions of the CppTransport [106, 76] and PyTransport [80, 77] codes that automate evaluation of inflationary correlation functions directly from low-energy effective action. Such automated methods are the most practical way to handle models whose numerical implementation is otherwise too laborious or error-prone, especially for calculation of three-point statistics. Some details of the improvements in the new versions of these codes are described below, but they will be discussed more completely in a forthcoming publication.

The underlying technology is an evolution of the ‘transport’ method already used to analyse the D3/D3 model by Dias et al. [107]. It has already been described in the literature [108, 109, 110, 111, 79, 75], and our implementation introduces no significant novelties compared to these treatments. Therefore we recapitulate only those properties relevant to our analysis or its interpretation. First, neither implementation makes use of the slow-roll approximation and therefore time dependence is treated exactly. However, in common with all other general-purpose frameworks for the calculation of inflationary correlation functions, the CppTransport and PyTransport implementations are valid only to tree-level. Here, ‘tree-level’ has its usual meaning in which a term at $n^{th}$ order in the loop expansion involves $n$ unrestricted momentum integrals. There are two types of loop in the Schwinger (or ‘in–in’) formalism appropriate for cosmological correlation functions [112, 113, 114].
The first type represent the familiar averages over virtual quanta that appear in ‘in–out’ amplitudes, and can be absorbed into a renormalization of masses, coupling constants and field amplitudes. The second type can be regarded as averages over unobserved physical particles, which may include decay products or particles generated from non-adiabatic evolution, including resonance [115, 116]. Momentum integrals of this type are a measure of back-reaction from these particle production processes. Any tree-level framework, including the transport method, is blind to this back-reaction. In this paper we simply assume there is no problematic back-reaction from particle production. However, see footnote 24 on p. 87.

What is included? In both two- and three-point functions we capture all effects from quadratic mixing between modes on superhorizon scales where momenta are soft compared to $H$ in the sense $k/(aH) \ll 1$, and even small off-diagonal terms in the mass matrix are relevant. In the traditional language of inflationary phenomenology these effects describe transfer of power between adiabatic and isocurvature modes. Meanwhile, in the three-point function we capture the effect of three-body interactions. These can loosely be regarded as describing processes in which a pair of particles are produced from the gravitational field, before one member of the pair decays into two daughter particles [117, 118]. At later times the three resulting particles are correlated due to their shared history.

We capture effects from any nontrivial mass spectrum, including modes that are much lighter ($m \ll H$), much heavier ($m \gg H$), or comparable to the Hubble scale ($m \sim H$). At horizon exit these effects can reduce the amplitude of fluctuations, or change the subtle interference effects imprinted in the three-point function—and higher-order correlations—from interaction between growing and decaying modes. They may also induce significant correlations, or anti-correlations, between the field degrees of freedom at horizon exit.

To detect the emergence of an adiabatic limit before the end of inflation we employ a technique based on tracking eigenvalues of the mass matrix [119, 120]. If an adiabatic limit is reached this implies the model is predictive without the need to specify details of a later reheating phase [121, 122, 123, 124, 125, 31, 119, 126, 127].

Because our interest lies in computation of observables, the CppTransport and PyTransport codes are only the front-end of a longer pipeline. Once the inflationary computation is complete, the two-point function is used as an initial condition for the CLASS Boltzmann code [128]. This enables us to generate custom predictions for the CMB angular

---

2A major advantage of the AdS/CFT computation used to obtain the D3-brane potential is that it accounts automatically for mixing between scales that is usually generated by averaging over virtual quanta.

3We use only the primordial two-point function (evaluated at the end of inflation) as an initial condition for the subsequent CMB calculation. We could equally well use this as an initial condition for the matter or galaxy power spectra, but we do not do so here because constraining the model from data is not the primary purpose of this paper and significant extra complexity is needed to describe galaxy bias. However, if desired, the flexibility of the CosmoSIS framework makes it simple to include more datasets in the likelihood calculation.

We will see below that computations of the primordial three-point function are still sufficiently expensive that we cannot routinely compute the shape of the bispectrum, for example to accurately model scale-dependent bias in the power spectrum, or to compute a full CMB bispectrum. We comment on the computational challenges in §2.2.1.2 and Appendix C below.
power spectra $C_\ell$ and hence the likelihood function. This is particularly important for the D3/D3 model because it frequently produces power spectra with significant scale dependence [107]. Summary statistics evaluated at a single scale—such as the scalar amplitude $A_s$ and spectral index $n_s$—are therefore misleading, and use of the full primordial power spectrum is required. The entire pipeline is controlled by the CosmoSIS parameter estimation framework [129], allowing it to be attached to a number of efficient general-purpose sampling algorithms. We collect values for observables built from the two- and three-point functions and use these to estimate distribution functions. Our focus is on general properties of perturbations produced by the model, whether or not their statistical character falls in an observationally viable window.

Using this pipeline we are able to study the distribution of the three-point correlation amplitude on representative ‘equilateral’ and ‘folded’ configurations (where $k_1 \sim k_2 \sim k_3$ and $k_1 \sim k_2 \sim k_3/2$, respectively, if $\mathbf{k}_1$, $\mathbf{k}_2$, $\mathbf{k}_3$ are the 3-momenta appearing in the three-point correlator). ‘Squeezed’ configurations, where one of the $k_i$ becomes significantly smaller than the other two, are substantially more expensive to simulate and we are not able to compute these for every realization. This is unfortunate because it is the squeezed correlation amplitude that can be measured most cleanly [130]. Instead, we study how the squeezed amplitudes correlate with the equilateral and folded ones by constructing a separate, smaller sample. From this we infer the behaviour of squeezed configurations in our full catalogue.

**Previous results.**—The brane inflation paradigm was introduced by Dvali & Tye [131], and elaborated into the concrete D3/D3 scenario by Burgess et al. [132] and Dvali et al. [133]. The branes carry opposite charges, and in early work the resulting Coulomb attraction was identified with the inflationary potential. Unfortunately, this proposal was not viable due to phenomenological difficulties.

Kachru et al. (‘KKLMMT’ [89], following earlier work by ‘KKLT’ [134]), showed that the Coulomb potential would be flattened by warping of the metric in the extradimensional space. Such warped-product geometries were already familiar from the Randall–Sundrum scenario [135, 136]. Although the flattened potential relieved most phenomenological problems, Kachru et al. demonstrated that it would receive significant corrections from effects due to moduli stabilization [89] and in particular that these would lift the inflaton mass to be of order $H$. This is a manifestation of the familiar $\eta$-problem of inflation. The conclusion is that successful D3/D3 inflation would have to be regarded as a ‘delicate’ accident caused by partial cancellation to produce a mass smaller than $H$.

An explicit computation of these corrections was given by Baumann et al. [137] for the case where stabilization is due to D7-branes wrapping four-cycles of the extradimensional space. The implications for D3/D3 inflation were summarized in Refs. [138, 139]. Their calculation showed that, by fine-tuning the cancellation between different effects, a small window existed for inflation to occur near an inflexion point of the D3-brane potential.
The approach used in these papers left open the question of what happens when moduli stabilization occurs by a more general mechanism. This issue was taken up by Baumann et al., first in a linearized analysis [140] and later including nonlinear effects [141, 142]. Their most developed method made use of the AdS/CFT correspondence to map operators in the D3-brane potential to the spectrum of non-normalizable perturbations of the warped conifold. These perturbations can be determined by harmonic analysis of the conifold base space, the coset space $T^{1,1}$, for which the necessary tools had already been assembled by Gubser [143] and Ceresole et al. [144, 145]. We review this method of constructing the brane potential in §2.1.

**Inflationary analysis.**—Each of these computations yields a prediction for the functional form of the D3-brane potential, parametrized by a number of mass scales $M_i$. These scales may be regarded as encoding ultraviolet information about the compactification that has been integrated out to produce the low-energy description. In our present state of ignorance they cannot be computed and must be estimated from observations.

If the low-energy effective action were to be expanded in a basis of local operators, the Wilson coefficient for each operator would be determined by the mass scales $M_i$. These coefficients would all have been regarded as independent by the methods of traditional effective field theory ('EFT'). For the D3 model, however, relationships inherited from the functional form of the brane potential imply that certain Wilson coefficients are correlated or even absent. It is these correlations that represent the gain in information from using an explicit ultraviolet completion in contrast to a traditional analysis using an ultraviolet-agnostic EFT. We will see below that the likelihood of inflation and its detailed predictions can depend on this pattern of correlations. This clearly illustrates the weak decoupling between inflation and the assumed ultraviolet model.

Agarwal et al. assessed these importance of these correlations between the low-energy coefficients by repeatedly drawing values for the $M_i$ from a specified distribution [146]—a strategy that had been introduced earlier by Easther, Peiris and collaborators [147, 148]. Assuming inflation would always begin from the same initial field configuration, Agarwal et al. were able to determine its likelihood as a function of the number of e-folds achieved. They also determined the distribution of the single-scale summary statistics $A_s$, $n_s$ and $r$ derived from the two-point function. Because the angular directions are typically heavy they argued that an effectively unique inflationary trajectory would often emerge, and used a single-field approximation based on this trajectory to estimate observables. This approximation does not capture multiple-field effects that transfer power between entropic modes and the curvature perturbation. It also does not account for the contribution of fields that are not light compared to the Hubble scale.

Agarwal et al. used an ensemble of $> 70,000,000$ realizations to study the homogeneous background, finding that the probability of more than 60 e-folds of inflation was small, of order $10^{-5}$ to $10^{-4}$. Two further ensembles were used to study observables: one with 4,900,000 realizations, of which 8,301 yielded more than 60 e-folds; and a second with
500,000 realizations, of which only 750 yielded more than 60 e-folds. The two ensembles differed in their truncation of the D3-brane potential, to be described in §2.1. Agarwal et al. concluded that the tensor–scalar ratio $r$ would typically be unobservable, and that the scalar spectral index $n_s$ fell roughly in the range $0.94 \lesssim n_s \lesssim 1.10$, with values in the WMAP7 range $n_s = 0.963 \pm 0.014$ (at $k = 0.002 \text{Mpc}^{-1}$ [149, 150]) coming from cases where the single-field treatment was likely to be acceptable. Their results did not depend strongly on the truncation.

In a small fraction of cases, Agarwal et al. observed abrupt transitions between different angular minima and speculated that these might generate significant non-Gaussianity from multiple-field effects [119, 120]. Our results demonstrate that this suggestion is essentially correct. Indeed, the amplitude of three-point correlations generated in this way can be surprisingly large, although trajectories that exhibit the effect are rare within our ensemble.

Dias et al. used a more sophisticated numerical scheme to compute observables [107], based on superhorizon ‘transport’ of the inflationary correlation functions [108, 109, 120, 66, 78]. This approach was a precursor of the technology we deploy in this paper. (The approach used here is more complete because it correctly accounts for subhorizon effects.) Their method correctly tracked transfer of power on superhorizon scales, including contributions from fields that were not light at the time of horizon exit. Like the scheme of Agarwal et al., it applied only to the two-point function.\footnote{Strictly this applies to v2 of the arXiv version of this paper, which includes an erratum to the published version (matching arXiv v1). Originally this paper contained an error from omission of the conifold metric, pointed out in Ref. [151], which caused all fields to be light at horizon exit. Based on this error the published version included a discussion of the $f_{NL}$ observable, but its conclusions were invalidated when the correct conifold metric was introduced in the erratum. We would like to thank Mafalda Dias and Jonathan Frazer for helpful correspondence, and for kindly sharing their Mathematica code.} As part of their analysis, Dias et al. attempted to quantify how many instances of inflation converged to an adiabatic limit. As explained above, when this occurs it implies that the model is predictive without specifying the details of a later reheating phase. Conversely, if an adiabatic limit is not reached, the final value of each observable may depend on the details of reheating [127].

Dias et al. used an ensemble with 564 realizations giving more than 60 e-folds of inflation. Within the statistical limits of their sample size, these results confirmed the conclusion of Agarwal et al. that $r$ would be negligible, and yielded a comparable distribution for $n_s$. In most cases, they found that an adiabatic limit would be reached during the inflationary phase.

Later, McAllister, Renaux-Petel & Xu studied the same model using a different numerical technique, finding 18,731 realizations that yielded at least 66 e-folds of inflation [151]. They found that 21% of their realizations were consistent with WMAP7 constraints on $n_s$ (see above). In agreement with Agarwal et al., these realizations typically exhibited a unique inflationary trajectory over the final 60 e-folds of inflation, making a single-field treatment sufficient. This usually occurs when inflation is of long duration, with multiple-field effects appearing only as transients at early times. Finally, based on an analytic approximation for the ‘quasi single-field’ regime [152, 153, 154], they suggested that three-point correlations...
on squeezed configurations could occasionally become large, with $|f_{NL}| \gtrsim 10$ in perhaps 0.07% of realizations. If it occurs, this form of non-Gaussianity has a very different origin to the rapid shifting between angular minima suggested by Agarwal et al. Unfortunately, detecting the presence of ‘quasi single-field’ effects is numerically expensive, and in this paper we are not yet able to form a definitive judgement regarding their occurrence.

A different approach was pursued by Hertog & Janssen [155], who studied the possibility of eternal inflation near the flat inflexion point that characterizes inflating potentials in the $D3$//$\overline{D3}$ model. This is very similar to the proposal of topological inflation [156, 157]; see Ref. [158] for an analysis of non-Gaussianity in related models. Hertog & Janssen computed observables in their scenario using a prior based on the no-boundary wavefunction proposal. Accordingly the observable distributions reported by these authors cannot be compared with those given here, although their suggestion $|f_{NL}| < 10^{-4}$ within their ensemble is notable.

**Organization of this paper.**—To build our primary catalogue required sampling more than 450,000,000 trajectories, of which over 90,000 yielded more than 60 e-folds of inflation. It will be explained in §2.2.3 that some of these are excluded due to concerns about representative sampling, leaving roughly 55,000 ‘safe’ trajectories for which observables can be computed. This is nearly three times the number of trajectories used by McAllister et al., and nearly 100 times the number used by Dias et al. The large sample size means we are able to characterize the distribution of each observable with reasonable accuracy. (However, we will see that there is evidence we still undersample away from the central region for some distributions.) The main obstruction to generating even larger ensembles is compute time. Running on a Haswell-era compute cluster, our production code required $\sim 95,000$ CPU hours to build the primary catalogue, and a further $\sim 40,000$ CPU hours to compute observables for each trajectory. We comment further on the resource requirements for the computation in Appendix C.

This primary catalogue is complemented by a number of ‘small’ catalogues, each comprising roughly 18,000 inflating trajectories, that are used to study the dependence of observables on various arbitrary choices made during construction of the model. These include the way the potential is truncated, the initial conditions for inflating trajectories, and the treatment of contributions to the brane potential from bulk fluxes (see §2.1.4). Each of these ‘small’ catalogues has similar size to the current best-in-class analysis reported in Ref. [151], giving us considerable statistical power when comparing distributions.

In §2.1 we review the construction of the D3-brane potential, paying particular attention to the harmonic modes on $T^{1,1}$. The elements of this discussion have all been given before, but are scattered across a number of papers. We collect the relevant formulae in a unified notation.

During the course of this work we discovered instances where inadvertent omissions meant that previous analyses of this model had not been described in sufficient detail to allow
replication. These relate to minor technical choices in the construction of the D3-brane potential or in specifying priors for the sampling procedure. To assist authors who wish to replicate our own analysis we have attempted to document the construction of the potential in sufficient detail to allow replication if desired. Our trajectory catalogues are available for download from the Zenodo open-access repository, and may be re-used under a permissive Creative Commons licence. Further, our computational pipelines are open source and published on GitHub. We would like to thank the authors of the previous studies for their ready assistance in relating our analysis to theirs.

The reader whose interest lies solely with the prediction of inflationary observables may wish to skip §2.1, which involves ideas from extradimensional compactifications in string theory, and return to it only to understand the relationship between parameters. For this purpose Tables 2.1 and 2.2 may be helpful. Conversely, readers who are already familiar with the detailed constructions of Refs. [141, 142] will not find any new material and may also wish to proceed directly to §2.2.

In §2.2 we describe our software stack and the numerical method used to compute observables. We document our choice of priors and initial conditions, and the precise sampling strategy we apply to build both the primary catalogue and the ‘small’ catalogues used for comparison. In §2.2.1.1 and §2.2.1.3 we explain how observables are computed within each pipeline, and in §2.2.1.2 we discuss general computational issues that arise within the transport framework irrespective of implementation. In §2.2.2 we describe our procedure for detecting an adiabatic limit at the end of inflation. Finally, in §2.2.3 we compare the distributions reported by each pipeline and develop a choice of cuts intended to ensure the integrity of our analysis. Imposing these cuts reduces our primary catalogue from (roughly) 90,000 to 55,000 trajectories, as explained above.

In §2.3 we study the distribution for each observable over the catalogues constructed in §2.2. §2.3.1 discusses the behaviour of trajectories at the level of the background. In §2.3.2.1 we consider observables derived from the two-point function, and in §2.3.2.2 we extend this to include information on three-point correlations on equilateral, folded, and (via a separate catalogue) squeezed configurations. We compare our distributions with results previously given in the literature. In §2.3.2.4 we discuss a population of rare trajectories that exhibit the abrupt transitions between angular minima observed by Agarwal et al., and show that these yield very large three-point correlations of ‘local’ shape. Finally, we conclude in §2.4. Three appendices summarize information tangential to the main discussion. provides information about the data deposit accompanying this paper. Appendix C gives more details on computational resource requirements. Appendix B summarizes the ‘transport’ computation of spectral indices, including new subleading terms intended to accelerate convergence.

Notation and conventions.—We work in natural units where \( c = \hbar = 1 \). The reduced Planck mass is defined by \( M_P = (8\pi G)^{-1/2} \) and is numerically equal to \( 2.435 \times 10^{18} \text{ GeV} \). Latin indices \( a, b, \ldots \), label coordinates in four-dimensional space, and indices \( A, B, \ldots \),
label coordinates in the six-dimensional compact space. The distributions $U(a, b)$ and $N(\mu, \sigma)$ are the uniform distribution with lower limit $a$ and upper limit $b$, and the normal distribution of mean $\mu$ and standard deviation $\sigma$, respectively.

The D3/$\overline{D3}$ model is very complicated—indeed, it may be the most complicated model for which inflationary observables have yet been computed. Its description entails a heavy overhead of notation. To assist readers who are unfamiliar with this maze of definitions we provide a glossary in Tables 2.1–2.2. Table 2.1 lists the parameters of the potential, together with their mass dimension, point of definition, and whether they are fixed, sampled, or derived quantities in our sampling procedure. Table 2.2 provides similar information for other relevant quantities that do not parametrize the potential.

In §2.1.2 and Appendix B we discuss the effective field theory for the D3-brane system and its fluctuations, for which we briefly summarize our conventions. The effective field theory comprises six scalar fields $X^A$ inherited from the extradimensional coordinates, with momenta $\pi^A = dX^A/dN$, where $N = \int H dt$ represents the accumulated e-folds of expansion. We collect the fields and momenta into a phase space coordinate $\mathcal{X}^A$ with index $A$. The corresponding fluctuations are $\delta \mathcal{X}^A$. 
Non-Gaussianity in D3-brane inflation

<table>
<thead>
<tr>
<th>parameter</th>
<th>mass dim.</th>
<th>definition</th>
<th>type</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r_{IR} = x_{IR}r_{UV}$</td>
<td>$-1$</td>
<td>IR limit of throat</td>
<td>p. 54 derived</td>
</tr>
<tr>
<td>$r_{UV} = \phi_{UV}/T_3^{1/2}$</td>
<td>$-1$</td>
<td>UV limit of throat</td>
<td>p. 54 derived</td>
</tr>
<tr>
<td>$T_3$</td>
<td>4</td>
<td>D3-brane tension</td>
<td>p. 54 fixed</td>
</tr>
<tr>
<td>$x_{IR}$</td>
<td>0</td>
<td>$r_{IR}$ in dimensionless $x$ coordinate</td>
<td>p. 56 fixed</td>
</tr>
<tr>
<td>$\phi_{UV}$</td>
<td>1</td>
<td>UV limit of throat in $\phi$ coordinate</td>
<td>p. 56 fixed</td>
</tr>
<tr>
<td>$Z = (2M_P/\phi_{UV})^2$</td>
<td>0</td>
<td>D3-brane charge</td>
<td>p. 56 derived</td>
</tr>
<tr>
<td>$a_0 = e^{A(r_{IR})}$</td>
<td>0</td>
<td>value of warp factor at IR limit of throat</td>
<td>p. 57 fixed</td>
</tr>
<tr>
<td>$\beta_s = e^\varphi$</td>
<td>0</td>
<td>string coupling</td>
<td>p. 56 not required</td>
</tr>
<tr>
<td>$D_0 = 2T_3a_0^4$</td>
<td>4</td>
<td>normalization of Coulomb potential</td>
<td>p. 57 derived</td>
</tr>
<tr>
<td>$V_0 = \alpha D_0$</td>
<td>4</td>
<td>energy density from distant supersymmetry breaking</td>
<td>p. 57 derived</td>
</tr>
<tr>
<td>$\mu = (V_0 + D_0)^{1/4}(\phi_{UV}/M_P)^{1/2}$</td>
<td>1</td>
<td>mass scale in $V_{\phi}$ and $V_{\phi}$</td>
<td>p. 57 derived</td>
</tr>
<tr>
<td>$C_{LM}$</td>
<td>0</td>
<td>Wilson coefficients for zero-model labelled by $LM$</td>
<td>p. 62 sampled</td>
</tr>
<tr>
<td>$D_{LM} = C_{LM} + C^*_{-L-M}$</td>
<td>0</td>
<td>combined Wilson coefficient for mode $LM$ if $L &gt; 0$</td>
<td>p. 64 derived</td>
</tr>
<tr>
<td>$\beta_{LM}$</td>
<td>0</td>
<td>Wilson coefficient for flux mode $LM$</td>
<td>p. 70 sampled</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>parameter</th>
<th>definition</th>
<th>type</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>relative magnitude of $V_0$ and $D_0$</td>
<td>—</td>
</tr>
<tr>
<td>$Q$</td>
<td>typical distance to nearest D7-brane stack</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>sampled$^b$</td>
<td>$U(-1,1)$</td>
</tr>
<tr>
<td></td>
<td>sampled$^c$</td>
<td>$U(0,0.04)$</td>
</tr>
</tbody>
</table>

$^a$ The initial conditions for the fields $\phi^A$ and their velocities $\dot{\phi}^A$ do not vary in our primary ensemble and do not appear in this table. See the discussion in §2.2.

$^b$ The $Q$ parameter was introduced by Agarwal et al. [146] [see Eq.(10) of this reference]. It effectively sets the normalization for the Wilson coefficients $C_{LM}$ and $C_{LM}^*$. This prior for $Q$ is selected with foreknowledge of the $Q$ posterior and a view to sampling efficiency. We will see in §2.3 that $Q$-values which frequently allow $N > 60$ e-folds of inflation are rather tightly clustered. We have verified that our final distributions for observables are unchanged if we select a weaker, broad prior for $Q$, except that the sampling process spends more time in regions of parameter space where $N > 60$ e-folds are unlikely. We use the uniform prior quoted here only in the CppTransport pipeline. The PyTransport pipeline instead uses a 'beta-prime' distribution $\beta'(\alpha, \beta)$ with shape parameters $\alpha = 4.16$, $\beta = 4.94$. See the discussion of the $Q$-prior in §2.2.3. Note that this $\alpha$ parameter is not the same as $\alpha = V_0/D_0$ appearing in the table.

Table 2.1: Glossary of parameters for the D3/$\overline{D3}$ model.
Table 2.2: Glossary of notation for the D3/$\overline{D3}$ model.

<table>
<thead>
<tr>
<th>quantity</th>
<th>definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T^{1,1}$</td>
<td>Romans manifold $SU(2) \times SU(2)/U(1)$; base of conifold</td>
</tr>
<tr>
<td>$G_{AB}$</td>
<td>conifold metric</td>
</tr>
<tr>
<td>$G_{AB}^{T^{1,1}}, ds_{T^{1,1}}^2$</td>
<td>metric on $T^{1,1}$</td>
</tr>
<tr>
<td>$A(r)$</td>
<td>Klebanov–Strassler warp factor</td>
</tr>
<tr>
<td>$H$</td>
<td>generator of $U(1)$ divisor in $T^{1,1}$</td>
</tr>
<tr>
<td>$K$</td>
<td>generator of $U(1)$ $R$-symmetry of $T^{1,1}$</td>
</tr>
<tr>
<td>$T_i$</td>
<td>generators of $SU(2)$</td>
</tr>
<tr>
<td>$\Psi = (\theta_1, \theta_2, \phi_1, \phi_2, \psi)$</td>
<td>coordinates on $T^{1,1}$</td>
</tr>
<tr>
<td>$r$</td>
<td>radial coordinate on the conifold</td>
</tr>
<tr>
<td>$C_4 = \alpha \omega$</td>
<td>Chern–Simons form coupling to D3-brane</td>
</tr>
<tr>
<td>$\omega$</td>
<td>volume form on D3-brane; $\omega = \star 1$</td>
</tr>
<tr>
<td>$\gamma_{ab}$</td>
<td>induced metric on D3-brane</td>
</tr>
<tr>
<td>$X^A$</td>
<td>embedding coordinates $X^A = X^A(x)$ of D3-brane</td>
</tr>
<tr>
<td>$\Phi_- = e^{4A} - \alpha$</td>
<td>supergravity field; AdS/CFT dual to D3-brane potential</td>
</tr>
<tr>
<td>$X = r/r_{UV}$</td>
<td>dimensionless D3-brane radial coordinate</td>
</tr>
<tr>
<td>$R_4$</td>
<td>4-dimensional Ricci scalar on D3-brane</td>
</tr>
<tr>
<td>$\Lambda$</td>
<td>3-form flux derived from $G_3$</td>
</tr>
<tr>
<td>$G_3$</td>
<td>3-form field of Type IIB supergravity</td>
</tr>
<tr>
<td>$V_\Phi$</td>
<td>Coulomb contribution to potential</td>
</tr>
<tr>
<td>$V_{\phi}$</td>
<td>‘complementary function’ part of potential</td>
</tr>
<tr>
<td>$V_{\phi}$</td>
<td>‘particular integral’ part of potential</td>
</tr>
<tr>
<td>$(\mu)$</td>
<td>label for representations of a Lie group $G$</td>
</tr>
<tr>
<td>$D_{mm'}^{(\mu)}(\beta)$</td>
<td>Wigner’s $D$-matrix in representation $(\mu)$</td>
</tr>
<tr>
<td>$d_{mm'}^{(\mu)}(\beta)$</td>
<td>Wigner’s little $d$ matrix in representation $(\mu)$</td>
</tr>
<tr>
<td>$L(g)$</td>
<td>coset representative of group element $g$</td>
</tr>
<tr>
<td>$L = (\ell_1, \ell_2, R)$</td>
<td>labels representation of isometry group $SU(2) \times SU(2) \times U(1)$</td>
</tr>
<tr>
<td>$M = (m_1, m_2)$</td>
<td>runs over representation space for representation $L$</td>
</tr>
<tr>
<td>$E_{LM}(\Psi)$</td>
<td>harmonic on $T^{1,1}$</td>
</tr>
<tr>
<td>$\Lambda_L$</td>
<td>eigenvalue of $E_{LM}$ in Laplacian $\nabla^2_{T^{1,1}}$</td>
</tr>
<tr>
<td>$J_{\ell,m,R}(\theta), J_{\ell,m,R}^\Omega(\theta)$</td>
<td>normalized $\theta$ modes on $T^{1,1}$</td>
</tr>
<tr>
<td>$\Delta(L) = -2 + \sqrt{1 + \Lambda_L}$</td>
<td>scaling dimension of zero-mode for representation $L$</td>
</tr>
<tr>
<td>$\Delta, \delta$</td>
<td>scaling dimensions unrelated to $L$</td>
</tr>
<tr>
<td>$\Theta(x,x'), g_L(r,r')$</td>
<td>Green’s function on conifold and eigensum decomposition $g_L$</td>
</tr>
<tr>
<td>$L, M, L', M'$</td>
<td>quantum numbers in Clebsch-Gordan expansion</td>
</tr>
<tr>
<td>$L, M$</td>
<td>$L = (L_1, L_2, R), M = (M_1, M_2)$</td>
</tr>
<tr>
<td>$A(L, L', M)$</td>
<td>unknown amplitude in Clebsch–Gordan expansion</td>
</tr>
</tbody>
</table>

*The $\alpha$ appearing in $C_4$ and $\Phi_-$ is the same, but not the same $\alpha$ that appears in Table 2.1.*
2.1 The D3/D3 model of inflation

The D3/D3 model has been studied extensively in the literature, to which we refer for a more complete account of its construction [131, 132, 89, 159, 137, 139, 138, 140, 141, 142]; see also Ref. [69] for a textbook description. In this section we recall only those details needed to fix notation or make our account self-contained. Our primary intention is to give an unambiguous specification of the D3-brane potential and its parametrization.

In §2.1.1 we recall the geometry of the conifold and explain how its four-dimensional low energy description can support an inflationary phase. In §2.1.2 we summarize the procedure of Baumann et al. for constructing the D3-brane potential [141, 142]. This depends on the details of harmonic analysis on the conifold, originally discussed by Gubser [143] and Ceresole et al. [144, 145], and summarized here from a different perspective in §2.1.3. These details are used to construct zero modes of the conifold Laplacian, which represent a class of contributions to the brane potential from ‘unsourced’ deformations of the conifold geometry. Finally, in §2.1.4 we give our prescription for a second class of ‘sourced’ deformations. Although these have been included in previous analyses, certain arbitrary choices are required to completely specify their contribution. In this section we document our choices in sufficient detail (at least in intention) to allow replication of our analysis.

2.1.1 The conifold geometry

In the model, a mobile D3-brane moves in an extradimensional space due to its mutual Coulomb attraction with a distant D3-brane. From the perspective of a four-dimensional observer, the displacement between the branes in each coordinate can be regarded as a four-dimensional scalar field. These fields will be nearly homogeneous if the branes are nearly parallel. The forces experienced by the brane as it moves in the extradimensional space can be summarized as an effective potential for the displacement in each coordinate.

The Klebanov–Strassler throat.—It was explained in §2 that the Coulomb force generates a potential that is too steep to support inflation if the extradimensional geometry is flat. To flatten the potential requires warping in the extradimensional space. A candidate geometry is the singular warped conifold studied by Klebanov & Strassler [160]. The ten-dimensional metric is a warped product,

\[
ds^2 = e^{2A} g_{ab} dx^a dx^b + e^{-2A} G_{AB} dX^A dX^B,
\]

where \( g_{ab} \) is the four-dimensional spacetime metric with coordinates \( x^a \) and \( G_{AB} \) is the metric on the transverse extradimensional space with coordinates \( X^A \). We take this to be a cone over the Romans space \( T^{1,1} = SU(2) \times SU(2)/U(1) \) [161],

\[
G_{AB} dX^A dX^B = dr^2 + r^2 ds^2_{T^{1,1}}.
\]
The geometry is supposed to be supported by a stack of $Z \gg 1$ D3-branes positioned at the tip. The base space $T^{1,1}$ is five-dimensional\(^5\) and carries the metric [162]

$$
\begin{align*}
\text{ds}_{T^{1,1}}^2 &= \frac{1}{9} \left( d\psi + \sum_{i=1}^{2} \cos \theta_i \, d\phi_i \right)^2 + \frac{1}{6} \sum_{i=1}^{2} \left( d\theta_i^2 + \sin^2 \theta_i \, d\phi_i^2 \right).
\end{align*}
$$

The coordinate ranges are $\theta_i \in [0, \pi]$, $\phi_i \in [0, 2\pi]$ and $\psi \in [0, 4\pi]$. The warp factor $A$ can be computed explicitly for the unperturbed conifold, but is not required for the inflationary analysis.

If the branes are exactly parallel, their displacement is uniquely labelled by the coordinate $r$. It runs from a minimum value $r_{\text{IR}}$, where the geometry is ‘deformed’ by attaching to a smooth cap that resolves the singularity at the tip of the cone [162], and extends in principle to arbitrarily large $r$. However, we usually imagine that the Klebanov–Strassler solution is cut off at some large value $r_{\text{UV}}$ and glued to a compact bulk space.\(^6\) In the field theory dual, small values of $r$ correspond to the infrared and large values correspond to the ultraviolet. The D$3$-brane is located in the infrared, where the warp factor minimizes its energy, and the mobile D3-brane is drawn from the ultraviolet towards the infrared end of the cone. Inflation ends when the mobile brane becomes sufficiently close to the D3 that a tachyonic instability develops and the brane pair dissolves into closed string modes.

In the four-dimensional field theory description this is a hybrid transition in which the inflaton is destabilized by a waterfall field. See Fig. 2.1.

**Kinetic and potential terms for the brane.**—The dynamics of the mobile D3-brane are determined by the action

$$
S = -T_3 \int d^4x \sqrt{-\det \gamma_{ab}} + T_3 \int d^4x \sqrt{-g} \, \alpha,
$$

where $x^a$ label coordinates on the brane and $\gamma_{ab}$ is its induced metric. There are two contributions. The first is the Nambu–Goto action. This computes the total energy of the brane, given by its tension $T_3$ integrated over its worldvolume. The second is a Chern–Simons term that couples the brane worldvolume to a four-form potential $C_4 = \alpha \omega$, where $\omega = \ast(1)$ is the volume form determined by $g_{ab}$.

---

\(^5\)For a description of coordinates on $T^{1,1}$, see Refs. [161, 162, 163, 164, 144, 145]. We follow Candelas & de la Ossa [162], especially §2 and Appendix A. $SU(2)$ is isomorphic to the three-sphere $S^3$, which itself is a Hopf fibration of $U(1)$ over $S^2$. The $U(1)$ can be regarded as parametrizing motion along a great circle of $S^3$. In the product $SU(2) \times SU(2)$ there are two $U(1)$ fibres corresponding to motion along great circles of the left and right $S^3$’s. We define the linear combinations $H = T_3 + T_1$ and $K = T_3 - T_1$, where the $T_1$ are generators of the left-hand copy of $SU(2)$ and $T_i$ are corresponding generators of the right-hand copy, and $T_3$ generate the $U(1)$ factors.

$T^{1,1}$ is the quotient $SU(2) \times SU(2)/U(1)_H$, obtained by identifying points that can be reached by a rotation generated by $H$. (It is part of a family of spaces $T^{p,q}$ obtained by generalizing $H$ to $H = pT_3 + qT_1$. To preserve supersymmetry we must choose $p = q = 1$ [161].) Therefore, in local Euler angles $\vartheta, \vartheta$ measured along each great circle, $T^{1,1}$ can be embedded as any hypersurface $\vartheta = \vartheta = \text{const}$ [143]. The coset representatives are labelled by the angular coordinate along this hypersurface. Accordingly, functions on $T^{1,1}$ should depend only on the coset label $\varrho = \vartheta$ and not $\varrho = \vartheta$.

\(^6\)The six extra dimensions must be compactified, unlike the Klebanov–Strassler solution, otherwise the effective Planck mass would be $\infty$ and there would be no dynamical gravity induced in the four-dimensional world.
Figure 2.1: Schematic representation of the extradimensional geometry. The throat is supported by a stack of D3-branes placed at its tip and is approximately described by the Klebanov–Strassler solution. It attaches in the ultraviolet \((x = 1, r = r_{UV})\) to an unknown bulk manifold. In the infrared \((r = r_{IR})\) a D3-brane draws the mobile brane down the throat due to their mutual Coulomb attraction.

We now drop the parallel approximation. Assuming the brane is embedded in the transverse dimensions at \(X^A = X^A(x^a)\), it follows that

\[
\gamma_{ab} = e^{2A} g_{ab} + e^{-2A} G_{AB} \partial_a X^A \partial_b X^B, \tag{2.5}
\]

where \(G_{AB}\) is the metric on the cone, Eq. (2.2). Therefore,

\[
S = -T_3 \int d^4 x \sqrt{-g} e^{4A} \sqrt{\det \left( \delta_{\ell} + e^{-4A} G_{AB} \partial_b X^A \partial_c X^B \right)} + T_3 \int d^4 x \sqrt{-g} \alpha. \tag{2.6}
\]

Assuming the brane is moving non-relativistically it suffices to work only up to quadratic order in derivatives. The zeroth order term from the determinant combines with the Chern–Simons term to generate a potential,

\[
V = T_3 \equiv T_3 \Phi_-, \quad \text{where} \quad \Phi_- = e^{4A} - \alpha. \tag{2.7}
\]

If the Klebanov–Strassler geometry is unperturbed—meaning that the infrared D3-brane is absent—then \(A = A(r)\) and \(\alpha = \alpha(r)\) depend only on the radial coordinate. Therefore \(V = V(r)\) also depends only on \(r\), and the angles \(\{\theta_1, \theta_2, \phi_1, \phi_2, \psi\}\) are flat directions. However, more careful analysis shows that in this case \(\alpha = e^{4A}\) and the potential vanishes \cite{165}. Generically, both \(A\) and \(\alpha\) will depend on \(r\) and the angles. This lifts the flat directions. Notice that the potential depends on the warp factor, which produces the flattening observed in Refs. \cite{134, 89}.

At second order in derivatives we obtain

\[
S = -\frac{T_3}{2} \int d^4 x \sqrt{-g} G_{AB} \partial^\alpha X^A \partial_{\alpha} X^B, \tag{2.8}
\]
in which the warp factor has cancelled. Eq. (2.8) is the kinetic term for a set of noncanonical four-dimensional scalar fields $X^A$ with kinetic mixing matrix $G_{AB}$ inherited from the cone (2.2). Therefore, if the potential can be chosen suitably, the $\phi^A$ may support a phase of slow-roll inflation. Predictions from inflationary models of this type were studied by Sasaki & Stewart [72]. The theory was developed up to three-point observables by a number of authors [166, 74, 167, 168, 169, 119, 170, 76].

Instead of $r$ we choose to work in terms of the coordinate $x = r/r_{UV}$ introduced in Ref. [146]. Its range is $x_{IR} < x \ll 1$, where $x_{IR} \equiv r_{IR}/r_{UV}$. The throat attaches to the compact bulk space in the region $x \sim 1$. Further, if we simplify the brane kinetic term by absorbing the tension $T_3$ into the metric, we find

$$G_{AB} dX^A dX^B \to r_{UV}^2 T_3 \left( dx^2 + x^2 ds_{T_{1,1}}^2 \right) \equiv \phi_{UV}^2 \left( dx^2 + x^2 ds_{T_{1,1}}^2 \right),$$

(2.9)

where we have defined $\phi_{UV} = r_{UV} T_3^{1/2}$. Note that the fields $\{x, \theta_1, \theta_2, \phi_1, \phi_2, \psi\}$ appearing in the four-dimensional effective action all have engineering dimension zero. To compensate, the metric $G_{AB}$ has engineering dimensions of $[\phi_{UV}^2] = [M^2]$.

Field range.—Baumann & McAllister argued that in a throat carrying D3-brane charge $Z \gg 1$, the field range would be bounded by $\phi_{UV} < 2M_P/Z^{1/2}$ [159]. Following Agarwal et al. [146] we generally take $\phi_{UV} = 10^{-1}$ and fix $T_3 = 10^{-2} M_P^2$. In §2.3.3 we briefly look at the effect of varying the field-range bound over the interval $10^{-1} < \phi_{UV} < 10^{-3}$.

2.1.2 The D3-brane potential

The remaining task is to enumerate permitted contributions to the D3-brane potential, Eq. (2.7). In the unperturbed Klebanov–Strassler geometry (without the infrared D3-brane) both $A = A(r)$ and $\alpha = \alpha(r)$ can be calculated explicitly [165], but as explained above this leads to a vanishing potential. The interpretation is that gravitational attraction between the mobile D3-brane and the D3-brane stack at the tip is balanced by repulsion due to their same-sign charges. This arrangement cannot support an inflationary epoch. To generate a nontrivial potential requires additional sources, so that we no longer expect exact cancellation.

How are we to determine the possible contributions? The general formula $V = T_3 \Phi_-$ given in (2.7) continues to apply, which reduces the problem to determination of $\Phi_-$. The supergravity field equations in the throat can be shown to require

$$\nabla^2 \Phi_- = R_4 + \frac{g_s}{96} |A|^2 + e^{-4A} |\nabla \Phi_-|^2 + \text{local terms},$$

(2.10)

where $g_s$ is the string coupling, $\nabla^2$ is the Laplacian on the conifold (2.2), and $R_4$ is the four-dimensional Ricci scalar. $A$ is a 3-form flux that depends on the 3-form field $G_3$ of type IIB supergravity; for details, see Refs. [140, 142, 69]. The ‘local terms’ represent localized contributions from the mobile brane and the D3-brane, which we now introduce.
Coulomb and mass terms.—First, the local terms generate a Coulomb attraction between the mobile brane and the anti-brane, with potential

\[ V_C(x) = D_0 \left( 1 - \frac{27}{64\pi^2} \frac{D_0}{\phi_{UV}^4} x^4 \right). \] (2.11)

The parameter \( D_0 \) is defined by \( D_0 \equiv 2T_3 a_0^4 \) [140], where \( a_0 \equiv e^{A(r_{IR})} \ll 1 \). It is the smallness of \( D_0 \), caused by warping of the conifold, that makes the potential sufficiently flat to inflate at modest values of \( x \).

Second, the leading effect of the Ricci term \( R_4 \) is to generate the operator \( R_4 \phi_{UV}^2 x^2/12 \). During inflation the background geometry is approximately de Sitter, for which \( R_4 = 12H^2 \). If we take the inflationary phase to be supported by the constant term in (2.11), possibly augmented by a second uplift \( V_0 \), then \( 3H^2 M_P^2 \approx V_0 + D_0 \). Here \( V_0 \) accounts for constant contributions that do not originate in the Coulomb interaction, which could include distant sources of supersymmetry breaking. In total this yields a mass term for \( x \) of the form

\[ V_M = \mu^4 x^2/3, \] (2.12)

where we have defined [146]

\[ \mu^4 = (V_0 + D_0) \left( \frac{\phi_{UV}}{M_P} \right)^2. \] (2.13)

Note that \( \mu \) has mass dimension \([M]\).

Deformations of the throat.—Third, the throat geometry (2.2) may be disturbed because of back-reaction from the passage of the brane. It may also be deformed by the suture between the throat and the compact bulk geometry. Any such disturbances will affect the dynamics of the brane and contribute to its effective potential.

In the vicinity of the suture an adequate description of \( \Phi_- \) will require boundary conditions that determine how information from the compact bulk is communicated to the ultraviolet end of the throat. This complicated structure for \( \Phi_- \) will generate a large number of operators in the effective theory whose Wilson coefficients depend sensitively on the ultraviolet data. In this region there is little hope of performing a realistic analysis of the model.

On the other hand, in the infrared region \( x \ll 1 \) we expect that renormalization group running will suppress most of these operators, leaving only a handful of the most relevant terms. In this region fewer Wilson coefficients must be specified, making the model significantly simpler to analyse. In particular, as described in §2, we can parametrize our ignorance of the ultraviolet boundary data by drawing these unknown Wilson coefficients from one or more suitable statistical distributions.

Any deformation of the throat must satisfy (2.10). We work to leading order in perturbations. (For details of the approximation scheme being used we refer to the original
The local terms and Ricci scalar generate only the additive contributions described above.\(^7\) The equation to be solved is therefore
\[
\nabla_0^2 \Phi = \frac{g_s}{96} |\Lambda|^2,
\]
where \(\nabla_0^2\) is the unperturbed conifold Laplacian. Notice that to solve (2.14) we do not need to know the behaviour of the remaining supergravity fields except for the dilaton that determines \(g_s\). Its general solution consists of a particular integral (or ‘flux term’) \(\Phi_F\) supported by the source term \(g_s |\Lambda|^2 / 96\) plus a complementary function (or ‘homogeneous term’) \(\Phi_H\) that satisfies the homogeneous equation. The complete potential is therefore
\[
V = V_\varepsilon + V_H + V_F + V_F',
\]
where \(V_F'\) and \(V_F\) are the potential terms generated by \(\Phi_H\) and \(\Phi_F\), respectively. We describe their construction in §§2.1.3–2.1.4 below.

### 2.1.3 Harmonic analysis on the conifold

**The Peter–Weyl theorem.—** Both \(V_F'\) and \(V_F\) can be analysed using the methods of harmonic analysis on Lie groups. According to the Peter–Weyl theorem, an orthonormal basis for square-integrable functions on a compact Lie group \(G\) is furnished by the matrix coefficients \(D^{(\mu)}_{mm'}(g)\), summed over all unitary irreducible representations \((\mu)\) \(^{[171]}\). These are defined to satisfy
\[
D^{(\mu)}_{mm'}(g) \equiv \langle m | \rho(g) | m' \rangle,
\]
where \(|m\rangle\) labels a basis for the representation \((\mu)\) and \(\rho\) is its representation map.

Specifically, for a square-integrable function \(\Phi\) and \(g \in G\), the Peter–Weyl theorem guarantees that \(\Phi\) can be represented as the sum
\[
\Phi(g) = \sum_{(\mu)} \sum_{mm'} c^{(\mu)}_{mm'} D^{(\mu)}_{mm'}(g),
\]
where \(c^{(\mu)}_{mm'}\) are coefficients depending on \(\Phi\). Hence, the \(D^{(\mu)}_{mm'}\) function as harmonics of \(G\) in a sense analogous to Fourier analysis. Observe that each representation occurs in (2.17a) with multiplicity equal to its dimension. Although we will not need this refinement, if \(\Phi\) transforms in an irreducible higher-dimensional representation \((\nu)\) of \(G\), in the sense
\[
\Phi^{(\nu)}(g' \cdot g) = \sum_{m'} D^{(\nu)}_{mm'}(g') \Phi^{(\nu)}_{m'}(g),
\]
then the expansion (2.17a) is shortened and only the \((\nu)\) representation is present, with

---

\(^7\)This is not true in general. As explained in Ref. \([142]\), the effect of the Ricci term is to dress each term in the potential with higher powers of \(x\). However, we will truncate the brane potential before the first of these dressed terms appears. See the discussion on p. 62. Therefore, for our analysis, it suffices to add the Coulomb term and mass term to the potential obtained from Eq. (2.14).
multiplicity one. For further details see Salam & Strathdee [172], who explained the application of (2.17a) to the typical case where \( \Phi \) is a supergravity field transforming in some nontrivial representation of the tangent space \( SO(1,3) \) symmetry.

Our interest lies in the case where \( \Phi \) is a spacetime scalar and \( G \) is the coset \( T^{1,1} = SU(2) \times SU(2)/U(1) \) described above. In this situation (2.17a) continues to apply, with

\[
\Phi[L(g)] = \sum_{(\mu)} \sum_{mm'} c^{(\mu)}_{mm'} D^{(\mu)}_{mm'}[L(g)],
\]

(2.17c)

where \( L(g) \) is the coset representative of \( g \).

**Application to \( T^{1,1} \).—** The unitary irreducible representations of \( SU(2) \) are labelled by their spin \( \ell \). The corresponding matrix coefficients are given by Wigner’s *Darstellung* or \( D \)-matrix,

\[
D^\ell_{mm'}(\phi, \theta, \vartheta) \equiv \langle \ell m | e^{-i\phi T_\phi} e^{-i\vartheta T_\vartheta} e^{-i\theta T_\theta} | \ell m' \rangle,
\]

(2.18)

where the \( T_i \) are generators of \( SU(2) \), \{\( \phi, \vartheta, \theta \)\} are corresponding Euler angles, and \(-\ell \leq m, m' \leq \ell \). Representations of \( SU(2) \times SU(2) \) are built from the tensor product of a pair of representations of spin \( \ell_1, \ell_2 \) associated with the left- and right-hand \( SU(2) \) factors. We distinguish these factors using the labels \( i = 1, 2 \), respectively. It follows that the corresponding harmonics are

\[
\varepsilon_{\ell_1,\ell_2}^{m_1,n_1,m_2,n_2}(\phi_1, \vartheta_1; \phi_2, \vartheta_2) \equiv N D^\ell_{m_1,n_1}(\phi_1, \vartheta_1) D^\ell_{m_2,n_2}(\phi_2, \vartheta_2)
\]

(2.19)

where the quantum numbers \( \ell_i, m_i, n_i \) satisfy the usual constraints for representations of \( SU(2) \). The prefactor \( N \) is a normalization to be determined.

In terms of these Euler angles, \( T^{1,1} \) can be embedded in \( SU(2) \times SU(2) \) as a hypersurface \( \Sigma \) satisfying \( \vartheta_1 + \vartheta_2 = \text{const.} \). (See the discussion in footnote 5 on p. 54, and the explicit discussion given by Gubser [143].) Eq. (2.17c) shows that the harmonics on \( T^{1,1} \) follow from (2.19) by restriction to suitable coset representatives, and therefore we must project out dependence on \( \vartheta_1 + \vartheta_2 \). The representations are labelled by \( \psi = (\vartheta_1 - \vartheta_2)/2 \), where \( 0 \leq \psi < 4\pi \). To obtain the correct projection, note that the Wigner \( D \)-matrix can be expressed in terms of the ‘little’ \( d \)-matrix, defined by

\[
D^\ell_{mn}(\phi, \theta, \vartheta) \equiv e^{-im\phi} d^\ell_{mn}(\theta) e^{-in\vartheta}.
\]

(2.20)

Therefore we must choose \( n_1 = -n_2 \). Note that \( d^\ell_{mn}(\theta) \) is real.

We write \( n_1 = -n_2 = R/2 \). After making a parity inversion on the \( i = 2 \) sphere, the metric on \( \Sigma \) can be brought to the canonical form (2.3). Using the transformation rule

---

8Ceresole et al. define a ‘scalar harmonic condition’, which in our language can be written \( m_1 = R/2, m_2 = -R/2 \) [144, 145]. Harmonics satisfying this conditions depend only on \( \Delta \phi = \phi_1 - \phi_2 \), and not \( \phi_1 \) or \( \phi_2 \) separately. They are ‘scalar’ in the sense that they are unchanged under the \( U(1)_H \) divisor of \( T^{1,1} \). Notice, however, that this condition is immaterial for the expansion of a typical spacetime scalar such as (2.17a), which contains representations of all dimensions, not just ‘scalar’ representations in the sense of Ceresole et al.
\[ d^\ell_{-m,-n}(\theta) = d^\ell_{mn}(-\theta) \]
we find that the harmonics can be written \(^8\)

\[ \mathcal{E}^{\ell_1,\ell_2}_{m_1,m_2,R/2}(\theta_1,\phi_1,\theta_2,\phi_2,\psi) = \mathcal{N}' \exp \left( \frac{R}{2} \psi + \sum_i m_i \phi_i \right) d^{\ell_1}_{m_1,R/2}(\theta_1) d^{\ell_2}_{m_2,R/2}(\theta_2), \tag{2.21} \]

where \(\mathcal{N}'\) is an adjusted normalization. To repeat, the properties of the quantum numbers follow from the selection rules for representations of \(SU(2)\), viz.,

- \(\ell_1\) and \(\ell_2\) are nonnegative and either both integers or both half-integers;
- \(m_1 \in \{-\ell_1,\ldots,\ell_1\}\) and \(m_2 \in \{-\ell_2,\ldots,\ell_2\}\); and
- \(R/2 \in \{-l,\ldots,l\}\) where \(l = \min(\ell_1,\ell_2)\).

This analysis clearly exhibits the Lie group structure underlying the harmonics. For practical calculations, however, we require explicit formulae for the \(\mathcal{E}_{LM}\). Here we borrow the economical notation of Baumann et al. \([137]\) in which the harmonics are distinguished by multi-indices \(L = (\ell_1,\ell_2,R)\) and \(M = (m_1,m_2)\). Specifically \(L\) labels the representation of the harmonic under the isometry group \(SU(2) \times SU(2) \times U(1)\), and \(M\) runs over the corresponding representation space.

**Explicit formulae.**—Explicit formulae for the \(\mathcal{E}_{LM}\) were obtained by Gubser using a direct analysis of their governing differential equations \([143]\). Later, a more extensive discussion was given by Ceresole et al. \([144, 145, 173]\), who used algebraic methods based on group theory \([172]\). The details were summarized by Baumann et al. \([137]\). Expressions for the zero-modes on the conifold were given in Ref. \([142]\); see also Ref. \([174]\). The expression (2.21) in terms of the little d-matrix was first given in Ref. \([175]\).\(^9\) Here we briefly collect these details in a unified notation.

The \(\mathcal{E}_{LM}\) are eigenfunctions of the Laplacian on \(T^{1,1}\) with eigenvalue \(\Lambda_L\),

\[ \nabla^2_{T^{1,1}} \mathcal{E}_{LM}(\Psi) = -\Lambda_L \mathcal{E}_{LM}(\Psi). \tag{2.22} \]

The eigenvalue spectrum depends on a sum of quadratic Casimir invariants for the representations specified by \(L\), but not the representation-space labels \(M\). It satisfies

\[ \Lambda_L = 6 \left( \ell_1(\ell_1 + 1) + \ell_2(\ell_2 + 1) - \frac{R^2}{8} \right). \tag{2.23} \]

\(^9\)Explicit formulae for the \(d^\ell_{mn}\) were given by Wigner \([176]\). Their generating function was computed by Schwinger \([177]\). The connexion between harmonics on \(T^{1,1}\) and Wigner’s little d-matrix was apparently not noticed prior to Ref. \([175]\).
The necessary nonsingular solutions for \( d_{m,R/2}^\ell(\theta) \) are\(^{10}\)

\[
\left( \frac{2\ell + 1}{2} \right)^{1/2} d_{m,R/2}^\ell(\theta) = J_{\ell,m,R}^T
\]

\[
\equiv N_{LM}^T (\sin \theta)^m \left( \cot \frac{\theta}{2} \right)^R 2F_1 \left( \begin{array}{c} -\ell + m, \ 1 + \ell + m \\ 1 + m - R/2 \end{array} | \sin^2 \frac{\theta}{2} \right),
\]

(2.24a)

if \( m \geq R/2 \), and

\[
\left( \frac{2\ell + 1}{2} \right)^{1/2} d_{m,R/2}^\ell(\theta) = J_{\ell,m,R}^\Omega
\]

\[
\equiv N_{LM}^\Omega (\sin \theta)^R \left( \cot \frac{\theta}{2} \right)^m 2F_1 \left( \begin{array}{c} -\ell + R/2, \ 1 + \ell + R/2 \\ 1 - m + R/2 \end{array} | \sin^2 \frac{\theta}{2} \right),
\]

(2.24b)

if \( m < R/2 \). As explained above, both solutions are real. Here, \( 2F_1(a, b; c | z) \) is the Gauss hypergeometric function, and we have introduced mode functions \( J_{\ell,m,R}(\theta) \) to match the notation of Ref. [137]. It follows from a Sturm–Liouville argument that the \( J_{\ell,m,R}(\theta) \) are orthogonal for fixed \( m, R \) in the measure \( \sin \theta \, d\theta \). Their normalization is fixed by adjusting \( N_{LM}^T, N_{LM}^\Omega \) so that

\[
\int_0^\pi \sin \theta \, J_{\ell,m,R}(\theta) J_{\ell',m,R}(\theta) \, d\theta = \delta_{\ell\ell'}.
\]

(2.25)

If we choose the normalization of the \( E_{LM} \) so that

\[
E_{LM}(\Psi) = J_{\ell_1,m_1,R}(\theta_1) J_{\ell_2,m_2,R}(\theta_2) \exp \left( m_1 \phi_1 + m_2 \phi_2 + \frac{R}{2} \psi \right),
\]

(2.26)

then the \( E_{LM} \) satisfy the larger orthogonality condition

\[
\int d^5 \Psi \left( - \det G_{T^{1,1}} \right)^{1/2} E_{LM}(\Psi) E_{LM'}^*(\Psi) = \delta_{LL'} \delta_{MM'},
\]

(2.27)

where \( \Psi \) stands schematically for the five angles on \( T^{1,1} \), and \( G_{T^{1,1}} \) is its metric.

**Zero-modes on the conifold.**—Each \( T^{1,1} \) harmonic can be promoted to a zero-mode of the conifold—that is, a solution of the homogeneous equation \( \nabla_0^2 \Phi_- = 0 \). A simple calculation shows that if \( E_{LM} \) is a harmonic on \( T^{1,1} \) with eigenvalue \( \Lambda_L \), then [137]

\[
f_{LM}(r, \Psi) = r^{\Delta(L)} E_{LM}(\Psi)
\]

(2.28)

is a zero-mode of \( \nabla_0^2 \), where

\[
\Delta(L) \equiv -2 \pm \sqrt{4 + \Lambda_L}.
\]

(2.29)

Therefore the complementary function for \( \Phi_- \) can be expressed as a linear combination of

---

\(^{10}\)These formulae match those quoted in Ref. [143]. The four cases given there can be related in pairs using an Euler transformation of the hypergeometric function.
these zero-modes. It will make a contribution to the brane potential of the form

\[ V_H(x, \Psi) = \mu^4 \sum_{LM} C_{LM} x^{\Delta(L)} \mathcal{E}_{LM}(\Psi) + \text{c.c.}, \tag{2.30} \]

where the scale \( \mu^4 \), defined in Eq. (2.13), has been inserted by hand to account for the factor of the tension \( T_3 \) appearing in the dictionary between \( \Phi^- \) and the brane potential \( V \). The \( C_{LM} \) are taken to be unknown (complex) Wilson coefficients, and \( \Delta(L) \) determines the radial scaling of each operator. In the region \( |x| \ll 1 \), away from the ultraviolet end of the throat, only a few operators of lowest scaling dimension will be relevant, as anticipated in the discussion above Eq. (2.14).

Although we would like to keep as many operators as possible, there are practical limitations. As we increase the number of terms that are retained, we incur corresponding costs from the automated symbolic manipulations carried out by the CppTransport and PyTransport platforms, and also in the numerical solution of the transport equations. We will see that it is already challenging to solve for the three-point function in a model of this complexity, so it is not realistic to attempt to retain operators of very high order. On the other hand, at a minimum, we would like to retain operators that contribute significantly to the effective cubic couplings. If these are large (as suggested by the parametric estimates given in Ref. [151]) they potentially source a large bispectrum from the quasi-single-field ‘QSFI’ mechanism [152, 153, 178, 154]. It follows that aggressive truncation of the potential risks serious misprediction for a key observable of the model.

In Table 2.3 we tabulate the lowest-lying zero modes of \( \nabla^2_0 \) with radial scaling dimensions that satisfy \( \Delta(L) \leq \Delta_{\text{max}} = 3.8 \). (This choice was made by Agarwal et al. and Dias et al. [146, 107]. McAllister et al. [175] did not give their truncation explicitly, but apparently used the same prescription.) We use this truncation in §2.2 to construct our primary statistical ensemble, giving sufficient headroom to capture large QSFI effects. At this level there are eleven contributing representations \( L = (\ell_1, \ell_2, R) \). However, it should be remembered that the number of modes is rather larger because the dimension of these representations lies between 3 (for \( \ell_1 = 1, \ell_2 = 0 \) and vice-versa) and 9 (for \( \ell_1 = \ell_2 = 1 \)), and as explained in §2.1.3 each representation contributes with multiplicity equal to its dimension. There is a unique constant mode with \( \ell_1 = \ell_2 = R = 0 \) that we omit; it is proportional to the unit operator and merely renormalizes the vacuum energy. Therefore its effect can be absorbed into \( V_0 \). (However, see Table 2.6.)

**Reality properties of the zero-modes.**—Because the ‘little’ \( d \)-matrices (or equivalently, the \( J \) mode functions) are real, complex conjugation simply reverses the sign of the labels \( R, m_1 \) and \( m_2 \). This follows from Eqs. (2.24a)–(2.24b) after making an Euler transformation of the hypergeometric function. Therefore modes with \( R < 0 \) in Table 2.3 are related to those with \( R > 0 \) by complex conjugation. A special case of this observation is that modes with \( R = m_1 = m_2 = 0 \) are purely real.

In the interest of clarity, we note that the sum in (2.30) is unrestricted and includes
Table 2.3: Scalar zero-modes of the conifold Laplacian $\nabla_0^2$. Tabulated values are the radial scaling dimension $\Delta(L)$; $SU(2) \times SU(2)$ representation labels $(\ell_1, \ell_2)$; the $U(1)$ representation label $R$; the mode normalization constant for $J_{\ell,m,R}(\theta)$; and the dimension of the representation. We limit inclusion to modes with lowest-lying radial scaling dimensions $\Delta(L) \leq 3.8$. This gives a total of 73 different modes, all of which occur with fixed multiplicities as described in §2.1.3.

<table>
<thead>
<tr>
<th>$\Delta(L)$</th>
<th>decimal scaling</th>
<th>$\ell_1$</th>
<th>$\ell_2$</th>
<th>$R$</th>
<th>normalization $\times \pi^{3/2}$</th>
<th>dimension</th>
</tr>
</thead>
<tbody>
<tr>
<td>3/2</td>
<td>1.5</td>
<td>1/2</td>
<td>1/2</td>
<td>-1</td>
<td>$3\sqrt{3}/2$</td>
<td>4</td>
</tr>
<tr>
<td>3/2</td>
<td>1.5</td>
<td>1/2</td>
<td>1/2</td>
<td>1</td>
<td>$3\sqrt{3}/4$</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>2.0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>$9/4$</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>2.0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>$9/4$</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>3.0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>$9\sqrt{3}/16$</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>3.0</td>
<td>1</td>
<td>1</td>
<td>-2</td>
<td>$9\sqrt{3}/4$</td>
<td>9</td>
</tr>
<tr>
<td>$2\sqrt{7} - 2$</td>
<td>3.2915</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>$9\sqrt{3}/4$</td>
<td>9</td>
</tr>
<tr>
<td>7/2</td>
<td>3.5</td>
<td>1/2</td>
<td>3/2</td>
<td>-1</td>
<td>$3\sqrt{6}/4$</td>
<td>8</td>
</tr>
<tr>
<td>7/2</td>
<td>3.5</td>
<td>1/2</td>
<td>3/2</td>
<td>1</td>
<td>$3\sqrt{6}/2$</td>
<td>8</td>
</tr>
<tr>
<td>7/2</td>
<td>3.5</td>
<td>3/2</td>
<td>1/2</td>
<td>-1</td>
<td>$3\sqrt{6}/2$</td>
<td>8</td>
</tr>
<tr>
<td>7/2</td>
<td>3.5</td>
<td>3/2</td>
<td>1/2</td>
<td>1</td>
<td>$3\sqrt{6}/4$</td>
<td>8</td>
</tr>
</tbody>
</table>

73
representations with both signs of $R$. Moreover, the coefficients $C_{LM}$ are taken to be independent for each set of quantum numbers $L, M$. First, consider a complex mode $E_{LM}$ for which at least one of the labels $R, m_1$ and $m_2$ is nonzero. For convenience we define $-L = (\ell_1, \ell_2, -R)$ and $-M = (-m_1, -m_2)$. The contribution of $E_{LM}$ to $V_{ \Phi }$ can be written as a sum of $E_{LM}$ and $E_{-L-M}$,

\[
V_{ \Phi } \supseteq \mu^4 x^{\Delta(L)} \left( C_{LM} E_{LM} + C_{LM}^* E_{LM}^* \right) = \mu^4 x^{\Delta(L)} \left( C_{LM} E_{LM} + C_{LM}^* E_{-L-M} \right),
\]

where the notation $\supseteq$ denotes that $V_{ \Phi }$ contains the indicated contribution, together with other contributions that have not been written. Meanwhile a similar relation holds for $E_{-L-M}$. In combination they yield

\[
V_{ \Phi } \supseteq 2 \mu^4 x^{\Delta(L)} \left( \text{Re}(D_{LM}) \text{Im}(E_{LM}) - \text{Im}(D_{LM}) \text{Re}(E_{LM}) \right),
\]

where $D_{LM} \equiv C_{LM} + C_{LM}^*$. It follows that we can equivalently restrict the sum in (2.30) to $L = (\ell_1, \ell_2, R)$ with $R > 0$ provided we adjust the summand to match (2.32). Later we will take the real and imaginary parts of $C_{LM}$ to be random variables drawn from some distribution $X$. It is important to note that the real and imaginary parts of $D_{LM}$ should then be drawn from the appropriate distributions for the sum and difference of two $X$-distributed random variables, respectively.

Second, consider a real mode with $R = m_1 = m_2 = 0$. This depends only on (at least one of) $\theta_1, \theta_2$, unless $\ell_1 = \ell_2 = 0$. In this case it is a constant and is excluded from the sum as described above. Therefore the contribution from this mode to $V_{ \Phi }$ is

\[
V_{ \Phi } \supseteq 2 \mu^4 x^{\Delta(L)} C_{LM} E_{LM}
\]

where now $C_{LM}$ is real with numerical value drawn from the distribution $X$.

### 2.1.4 Flux contributions

The final step is to include the ‘particular integral’ for (2.14)—that is, the perturbation to $\Phi_-$ sourced by the square of the 3-form flux $|\Lambda|^2$. The allowed contributions to $\Lambda$ were enumerated by Baumann et al. [142] and fall into three distinct ‘series’, distinguished by the recipe for building the 3-form $\Lambda$ from a seed zero-mode drawn from Table 2.3. Depending on the details of the recipe, the resulting fluxes may exhibit an adjusted radial scaling dimension $\Delta$ or an adjusted $U(1)$ quantum number $R$. We will not require explicit formulae for the fluxes, and therefore refer to the literature for details of their construction.

**Flux contributions to $\Phi_-$**.—In Tables 2.4–2.6 we tabulate the required seed representations, and adjusted radial scaling dimensions, associated with these flux series, labelled Series I, II and III in the notation of Baumann et al. [142]. We limit attention to representations for which the adjusted scaling dimension $\Delta$ satisfies $2 < \Delta \leq 5.8$. At the upper
limit, we will see later that fluxes with $\Delta \leq 5.8$ are sufficient to capture all contributions to $\Phi_-$ with radial scaling dimension $\delta \leq \Delta_{\text{max}} = 3.8$. Notice that in this section we carefully distinguish the adjusted radial scaling dimension $\Delta$ of the fluxes, and the dimension $\delta$ of the contribution that is ultimately produced in the potential. At the lower limit, fluxes with $\Delta \leq 2$ do not couple to a probe D3 brane [142] and can be discarded. When summed over all three series there are 197 flux modes. Clearly the resulting D3-brane potential is very complicated.

The possible contributions to $|\Lambda|^2$ involve combinations of any two fluxes drawn from Tables 2.4–2.6, with the proviso that two chiral modes can combine only if they belong to the same flux series [142]. A flux mode is said to be chiral if and only if its seed scalar mode is chiral in the sense $\ell_1 = \ell_2 = R/2$ [142].

The contribution to $\Phi_-$ from any pair of fluxes can be found using the method of Green’s functions. Specifically,

$$\Phi_-(x) = \frac{g_s}{96} \int d^6 y \left( - \det G_{AB} \right)^{1/2} \mathcal{G}(x, y)|\Lambda(y)|^2,$$

(2.34)

where $G_{AB}$ is the metric (2.2) on the conifold and $\mathcal{G}(x, y)$ is the corresponding Green’s function obtained in Ref. [137]. It has the spectral representation

$$\mathcal{G}(x, x') = \sum_{LM} E_{LM}(\Psi)E^{*}_{LM}(\Psi')g_L(r, r'),$$

(2.35)

where $x = (r, \Psi)$, $x' = (r', \Psi')$ are coordinates on the conifold, and $\Psi$, $\Psi'$ represent the angles on $T^{1,1}$. The kernel $g_L(r, r')$ satisfies [137]

$$g_L(r, r') = -\frac{1}{2\Delta(L) + 1} \begin{cases} (r')^{-4} (r/r')^{\Delta(L)} & r_{\text{IR}} \lesssim r \lesssim r' \quad \text{for} \quad r' < r' \lesssim r_{\text{UV}}^\Delta, \\ r^{-4} (r'/r)^{\Delta(L)} & r' \lesssim r \lesssim r_{\text{UV}} \end{cases}.$$  

(2.36)

As explained above, some of the fluxes listed in Tables 2.4–2.6 have modified quantum numbers because they combine with other ingredients. For example (now passing to a Kähler description), some flux series involve the holomorphic 3-form $\Omega_{abc} = \epsilon_{abc}$, where the indices $a$, $b$, $c$ run over complex coordinates on the conifold, $\epsilon$ is the Levi–Civita tensor, and $q^* q = (- \det G_{AB})^{1/2}$ is the determinant of the Kähler metric. However, in $|\Lambda|^2$ these factors cancel with contractions involving the inverse metric. Therefore the angular terms in (2.34) involve only a product of the seed modes appearing in these tables.

**Radial profile.**—The integral in (2.34) factorizes into an integral over the radius $r$ of the cone and an integral over the angles on $T^{1,1}$. First, consider the radial integral. For the fluxes described in Tables 2.4–2.6, radial dependence arises only from their scaling dimensions. Accordingly, given two fluxes $\Lambda_1$, $\Lambda_2$ and a fixed representation $L$ drawn from the sum in (2.35), the Green’s function produces a radial profile

$$\text{radial profile} \propto \int_{r_{\text{IR}}}^{r_{\text{UV}}} dr' \frac{dr'}{r'} g_L(r, r') \left( \frac{r'}{r_{\text{UV}}} \right)^{\Delta_1 + \Delta_2},$$

(2.37)
Table 2.4: Series I fluxes. Tabulated values are the adjusted radial scaling dimension $\Delta = 1 + \Delta_f$, where $f$ is the scalar seed mode; the $SU(2) \times SU(2) \times U(1)$ quantum numbers $\ell_1, \ell_2, R$ for $f$; dimension of the representation; and whether the flux is chiral, defined to mean $\ell_1 = \ell_2 = R/2$ for the seed mode. We include the 109 modes with radial scaling dimensions that satisfy $2 < \Delta \leq 5.8$.

<table>
<thead>
<tr>
<th>$\Delta$</th>
<th>decimal scaling</th>
<th>$\ell_1$</th>
<th>$\ell_2$</th>
<th>$R$</th>
<th>dimension</th>
<th>type</th>
</tr>
</thead>
<tbody>
<tr>
<td>5/2</td>
<td>2.5</td>
<td>1/2</td>
<td>1/2</td>
<td>−1</td>
<td>4</td>
<td>non-chiral</td>
</tr>
<tr>
<td>5/2</td>
<td>2.5</td>
<td>1/2</td>
<td>1/2</td>
<td>1</td>
<td>4</td>
<td>chiral</td>
</tr>
<tr>
<td>4°</td>
<td>4.0</td>
<td>1</td>
<td>1</td>
<td>−2</td>
<td>9</td>
<td>non-chiral</td>
</tr>
<tr>
<td>4</td>
<td>4.0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>9</td>
<td>chiral</td>
</tr>
<tr>
<td>$\sqrt{28} - 1$</td>
<td>4.2915</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>9</td>
<td>non-chiral</td>
</tr>
<tr>
<td>9/2</td>
<td>4.5</td>
<td>1/2</td>
<td>3/2</td>
<td>−1</td>
<td>8</td>
<td>non-chiral</td>
</tr>
<tr>
<td>9/2</td>
<td>4.5</td>
<td>1/2</td>
<td>3/2</td>
<td>1</td>
<td>8</td>
<td>non-chiral</td>
</tr>
<tr>
<td>9/2</td>
<td>4.5</td>
<td>3/2</td>
<td>1/2</td>
<td>−1</td>
<td>8</td>
<td>non-chiral</td>
</tr>
<tr>
<td>9/2</td>
<td>4.5</td>
<td>3/2</td>
<td>1/2</td>
<td>1</td>
<td>8</td>
<td>non-chiral</td>
</tr>
<tr>
<td>$2\sqrt{10} - 1$</td>
<td>5.3246</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>non-chiral</td>
</tr>
<tr>
<td>$2\sqrt{10} - 1$</td>
<td>5.3246</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>5</td>
<td>non-chiral</td>
</tr>
<tr>
<td>11/2</td>
<td>5.5</td>
<td>3/2</td>
<td>3/2</td>
<td>−3</td>
<td>16</td>
<td>non-chiral</td>
</tr>
<tr>
<td>11/2</td>
<td>5.5</td>
<td>3/2</td>
<td>3/2</td>
<td>3</td>
<td>16</td>
<td>chiral</td>
</tr>
</tbody>
</table>

° The representations with $(\ell_1, \ell_2, R) = (1, 0, 0), (0, 1, 0)$ are absent, even though they appear in Table 2.3. These modes were excluded by Baumann et al. [142] without comment, apparently because they are projected out of the spectrum by the field equation for $G_3$. These seed modes do appear for the Series II and Series III fluxes listed in Tables 2.5 and 2.6. In Ref. [107] they were accidentally excluded from Series II and III. In principle this could influence outcomes from the model, but see footnote 12 on p. 69. We would like to thank Mafalda Dias for very helpful correspondence on these issues.
Table 2.5: Series II fluxes. Tabulated values are the adjusted radial scaling dimension Δ = 2 + Δf, where f is the scalar seed mode; the SU(2) × SU(2) × U(1) quantum numbers ℓ1, ℓ2, R for f; dimension of the representation; and whether the flux is chiral, defined to mean ℓ1 = ℓ2 = R/2 for the seed mode. We include the 73 modes with radial scaling dimensions that satisfy 2 < Δ ≤ 5.8.

<table>
<thead>
<tr>
<th>Δ</th>
<th>decimal scaling</th>
<th>ℓ1</th>
<th>ℓ2</th>
<th>R</th>
<th>dimension</th>
<th>type</th>
</tr>
</thead>
<tbody>
<tr>
<td>7/2</td>
<td>3.5</td>
<td>1/2</td>
<td>1/2</td>
<td>-1</td>
<td>4</td>
<td>non-chiral</td>
</tr>
<tr>
<td>4</td>
<td>4.0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>non-chiral</td>
</tr>
<tr>
<td>5</td>
<td>5.0</td>
<td>1</td>
<td>1</td>
<td>-2</td>
<td>9</td>
<td>non-chiral</td>
</tr>
<tr>
<td>5</td>
<td>5.0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>9</td>
<td>chiral</td>
</tr>
<tr>
<td>2√7</td>
<td>5.2915</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>9</td>
<td>non-chiral</td>
</tr>
<tr>
<td>11/2</td>
<td>5.5</td>
<td>1/2</td>
<td>3/2</td>
<td>-1</td>
<td>8</td>
<td>non-chiral</td>
</tr>
<tr>
<td>11/2</td>
<td>5.5</td>
<td>1/2</td>
<td>3/2</td>
<td>1</td>
<td>8</td>
<td>non-chiral</td>
</tr>
<tr>
<td>11/2</td>
<td>5.5</td>
<td>3/2</td>
<td>1/2</td>
<td>-1</td>
<td>8</td>
<td>non-chiral</td>
</tr>
<tr>
<td>11/2</td>
<td>5.5</td>
<td>3/2</td>
<td>1/2</td>
<td>1</td>
<td>8</td>
<td>non-chiral</td>
</tr>
</tbody>
</table>

Table 2.6: Series III fluxes. Tabulated values are the adjusted radial scaling dimension Δ = 3 + Δf, where f is the scalar seed mode; the SU(2) × SU(2) × U(1) quantum numbers ℓ1, ℓ2, R for f; dimension of the representation; and whether the flux is chiral, defined to mean ℓ1 = ℓ2 = R/2 for the seed mode. We include the 15 modes with radial scaling dimensions that satisfy 2 < Δ ≤ 5.8.

<table>
<thead>
<tr>
<th>Δ</th>
<th>decimal scaling</th>
<th>ℓ1</th>
<th>ℓ2</th>
<th>R</th>
<th>dimension</th>
<th>type</th>
</tr>
</thead>
<tbody>
<tr>
<td>3a</td>
<td>3.0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>chiral</td>
</tr>
<tr>
<td>9/2</td>
<td>4.5</td>
<td>1/2</td>
<td>1/2</td>
<td>1</td>
<td>4</td>
<td>chiral</td>
</tr>
<tr>
<td>9/2</td>
<td>4.5</td>
<td>1/2</td>
<td>1/2</td>
<td>-1</td>
<td>4</td>
<td>non-chiral</td>
</tr>
<tr>
<td>5</td>
<td>5.0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>non-chiral</td>
</tr>
<tr>
<td>5</td>
<td>5.0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>non-chiral</td>
</tr>
</tbody>
</table>

*Considered as a zero-mode of the scalar Laplacian ∇2, this mode is a constant. It was excluded from Table 2.3 because it does not contribute to V_H as explained in the main text. It also does not appear in the Series I or II fluxes, Tables 2.4–2.5, because these fluxes are all built from derivatives of the seed f. However, Series III fluxes include contributions from f without differentiation, so this mode can yield a nontrivial flux [142].
where the factor $1/r'$ is produced by combining $(r')^5$ from the Jacobian $(-\det G_{AB})^{1/2}$ and $(r')^{-6}$ from three copies of the inverse metric for the contractions implied by $|\Lambda|^2$. There is an overall constant of proportionality that we do not write explicitly. The result is

$$\text{radial profile} \propto \frac{1}{r_{\text{UV}}} \left( \alpha x^{\Delta_1+\Delta_2-4} + \beta x^{\Delta(L)} + \gamma x^{-4-\Delta(L)} \left( \frac{r_{\text{IR}}}{r_{\text{UV}}} \right)^{\Delta(L)+\Delta_1+\Delta_2} \right),$$

where $x = r/r_{\text{UV}}$ as above, and $\alpha$, $\beta$, $\gamma$ are roughly $O(1)$ numerical coefficients. The term involving the infrared regulator $r_{\text{IR}}$ is small provided $r_{\text{IR}}/r_{\text{UV}} \ll 1$ and formally vanishes in the limit $r_{\text{IR}} \to 0$. Therefore the integral does not accumulate large contributions from the region $r \sim r_{\text{IR}}$ where we need a precise resolution of the conifold singularity. We assume this term is negligible, and it will be dropped in the following discussion.

The remaining terms source radial profiles $\sim x^{\Delta(L)}$ and $\sim x^{\Delta_1+\Delta_2-4}$. The $x^{\Delta(L)}$ profile reproduces the radial scaling dimension associated with the scalar zero-modes of Table 2.3. This term will modify the coefficients associated with the $L$-representation in $V_{\Psi'}$, Eq. (2.30). Fortunately, this is harmless: our premise is that we cannot predict these coefficients, which already depend on ultraviolet data. Combining two unknown coefficients merely yields another unknown coefficient. The other term, scaling like $x^{\Delta_1+\Delta_2-4}$, is new. It will produce contributions to the potential involving modes in the $L$-representation, but with a radial scaling dimension $\Delta_{12} = \Delta_1 + \Delta_2 - 4$ different to $\Delta(L)$. Note that $\Delta_{12}$ is guaranteed to be positive because $\Delta > 2$ for all flux modes that participate in the cross product. If this dimension falls below the truncation point then such terms should be retained.\footnote{In Tables 2.4–2.6 we retained terms with $2 < \Delta \leq 5.8$. Inspection of the formula for $\Delta_{12}$ shows that to capture contributions to $\Phi_-$ with $\Delta \leq 3.8$ it is sufficient to consider fluxes with scaling dimension in this range.}

**Angular terms.**—Now consider the angular part of (2.34). We have already explained that factors coming from copies of the inverse metric in the contraction $|\Lambda|^2$ cancel with normalization adjustments in the individual fluxes. Therefore the integrand involves only the combination

$$\text{angular part} \propto \sum_{LM} \mathcal{E}_{LM}(\Psi) \int d^5\Psi' \left( -\det G_{T=1,1} \right)^{1/2} \mathcal{E}^*_{LM}(\Psi') \mathcal{E}^*_{L'M'}(\Psi') \mathcal{E}^*_{L''M''}(\Psi').$$

The modes $\mathcal{E}_{L'M'}$ and $\mathcal{E}_{L''M''}$ represent the quantum numbers of the two flux modes contributing to $|\Lambda|^2$. The two modes with quantum numbers $LM$ contribute to the sum in the spectral representation of the Green’s function, Eq. (2.35).

The complex conjugation on $\mathcal{E}_{L''M''}$ can be dropped without loss of generality, because this merely reverses the labels $L'' \to -L''$ and $M'' \to -M''$. To build $\Phi_-$ we will sum $(L', M')$ and $(L'', M'')$ over all entries in Tables 2.4–2.6, so this reversal is immaterial. Moreover, since $\mathcal{E}$ transforms as a pair of $SU(2)$ representations, the tensor product $\mathcal{E}_{L_1M_1} \mathcal{E}_{L_2M_2}$ can be decomposed as a direct sum of similar representations. This follows from the
same property of SU(2), expressed via Clebsch–Gordan coefficients. Specifically, Wigner’s ‘little’ $d$-matrix satisfies
\[ d^\ell_m(\beta)d^\ell'_{m'}(\beta) = \sum_{L=|\ell-\ell'|}^{\ell+\ell'} \langle \ell m, \ell' m' | L M \rangle \langle \ell n, \ell' n' | L N \rangle d^L_{MN}(\beta), \] (2.40)
where $M = m + m'$ and $N = n + n'$. Here, $\langle \ell m, \ell' m' | L M \rangle$ is an ordinary Clebsch–Gordan coefficient. The $J_{\ell_m,\ell}(\theta)$ mode functions used to build the $E_{LM}$ harmonics are related to $d^\ell_m$ via (2.24a)–(2.24b). Therefore, recalling $L = (\ell_1, \ell_2, R), \: L' = (\ell'_1, \ell'_2, R')$, $M = (m_1, m_2)$ and $M' = (m'_1, m'_2)$, we conclude
\[
\mathcal{E}_{LM}(\Psi) \mathcal{E}_{L'M'}(\Psi) = \sum_{\mathcal{L}_1=|\ell_1-\ell'_1|}^{\ell_1+\ell'_1} \sum_{\mathcal{L}_2=|\ell_2-\ell'_2|}^{\ell_2+\ell'_2} \sqrt{\frac{(2\ell_1+1)(2\ell'_1+1)(2\ell_2+1)(2\ell'_2+1)}{2(2\mathcal{L}_1+1)2(2\mathcal{L}_2+1)}} \times \langle \ell_1 m_1, \ell'_1 m'_1 | \mathcal{L}_1 \mathcal{M}_1 \rangle \langle \ell_1 R/2, \ell'_1 R'/2 | \mathcal{L}_1 / 2 \rangle \\
\times \langle \ell_2 m_2, \ell'_2 m'_2 | \mathcal{L}_2 \mathcal{M}_2 \rangle \langle \ell_2 R/2, \ell'_2 R'/2 | \mathcal{L}_2 / 2 \rangle \mathcal{E}_{LM}(\Psi), \] (2.41)
where $\mathcal{M}_1 = m_1 + m'_1, \: \mathcal{M}_2 = m_2 + m'_2, \: \mathcal{R} = R + R', \: \mathcal{L} = (\mathcal{L}_1, \mathcal{L}_2, \mathcal{R})$ and $\mathcal{M} = (\mathcal{M}_1, \mathcal{M}_2)$.\(^{12}\)

Substitution of (2.41) in (2.39) and use of the completeness relation (2.27) shows that the integral $d^\Psi \Psi'$ collapses to $\delta_{\mathcal{L} \mathcal{L'}} \delta_{\mathcal{M} \mathcal{M}'}$. For fixed $(L', M')$, $(L'', M'')$, the angular part therefore reproduces the right-hand side of (2.41) after suitable relabelling of indices. To summarize, consider the cross product between any two fluxes $\Lambda, \Lambda'$ drawn from Tables 2.4–2.6, with labels $(L, M)$ and $(L', M')$ and (adjusted) radial scaling dimensions $\Delta, \Delta'$. Unless both fluxes are chiral, their cross product makes a contribution to the brane potential, via the particular integral for $\Phi_{-}$, of the form
\[
\Lambda \times \Lambda' \rightarrow \frac{g_s}{96} \mu^4 \sum_{\mathcal{L}_1=|\ell_1-\ell'_1|}^{\ell_1+\ell'_1} \sum_{\mathcal{L}_2=|\ell_2-\ell'_2|}^{\ell_2+\ell'_2} \mathcal{A}(L, L', \mathcal{L}) \alpha x^{\Delta+\Delta'-4} \\
\times \langle \ell_1 m_1, \ell'_1 m'_1 | \mathcal{L}_1 \mathcal{M}_1 \rangle \langle \ell_1 R/2, \ell'_1 R'/2 | \mathcal{L}_1 / 2 \rangle \\
\times \langle \ell_2 m_2, \ell'_2 m'_2 | \mathcal{L}_2 \mathcal{M}_2 \rangle \langle \ell_2 R/2, \ell'_2 R'/2 | \mathcal{L}_2 / 2 \rangle \mathcal{E}_{LM}(\Psi). \] (2.42)

As explained above, we have dropped corrections from the infrared end of the conifold and from the term scaling like $x^{\Delta(\mathcal{L})}$ which is already included in $V_{\mathcal{M}}$.

If both fluxes are chiral then this expression applies if $\Lambda$ and $\Lambda'$ are drawn from the same

\(^{12}\)Ref. [107] used a direct numerical evaluation of (2.39) to compute the re-expansion of flux cross-products into $E_{LM}$ harmonics, but their implementation inadvertently neglected the angular Jacobian ($-\det G_{T^{-1}})^{1/2}$. This will slightly change numerical values appearing in the re-expansion, and it may also change the selection rules that couple the quantum numbers $(L, M)$, $(L', M')$ and $(L'', M'')$. Nevertheless, based on the numerical evidence to be discussed in §2.3.3 below it seems possible that this will not significantly influence the final distribution of observables reported by these authors.
series; if not, their contribution should be set to zero. The total potential $V_\mathcal{F}$ should be obtained by summing (2.42) over all fluxes $\Lambda, \Lambda'$. In this expression $\mathcal{A}(L, L', L)$ is an unknown amplitude that absorbs the unknown constants of proportionality in (2.37) and (2.39). It depends on the recipes used to construct $\Lambda$ and $\Lambda'$ from their seed zero-modes, and also the amplitude with which these ‘building block’ fluxes appear in the supergravity solution for $G_3$. It also absorbs the normalization factor that appears under the square-root in Eq. (2.41).

The final result is very complicated, and depends on constants such as $\mathcal{A}(L, L', L)$ that we cannot predict. To use it in a practical analysis one must make a number of largely arbitrary choices. Unfortunately, prior analyses of this model have generally not documented the choices made to convert Eq. (2.42) to a practical expression for the potential. In §2.3.3 we discuss numerical experiments in which the flux-sourced contributions described in this section are dropped, and find that this significantly affects the mass spectrum. The impact on observables is more modest but not negligible, and therefore a precise specification of $V_\mathcal{F}$ is necessary.

Our choices are as follows. First, we do not attempt to model the numerical coefficient $g_s \alpha_\mathcal{A}/96$ that normalizes Eq. (2.42). We collect these numerical factors into a single statistical Wilson coefficient $C_{LM}$ whose real and imaginary parts are assumed to be drawn from the same distributions that characterize the coefficients $C_{LM}$ in Eq. (2.30); see Table 2.1. Second, we do track the numerical value of the Clebsch–Gordan factors. If the same mode $x^{\Delta + \Delta' - 4} E_{LM}$ is produced from more than one cross-product of the fluxes in Tables 2.4–2.6, we add their numerical coefficients coherently to produce a single numerical prefactor. An alternative would be to treat each occurrence as an independent random variable, rather than add the amplitudes coherently. Yet another choice would be to model the Clebsch–Gordan coefficients as a Kronecker-$\delta$, equal to zero if the Clebsch–Gordan factors give zero and unity otherwise. Our procedure is intended to model, at least approximately, cases where the Clebsch–Gordan coefficients are unusually large or small, without causing a proliferation of parameters that unnecessarily enlarge the sample space. In practice the Clebsch–Gordan values typically do not vary significantly for the range of quantum numbers we are using and are almost always $O(1)$.

Notwithstanding the foregoing discussion, the analysis discussed in §2.3.3 does offer hope that the precise procedure used to model the amplitude of individual contributions to the potential will not radically alter the final distribution of observables. A detailed

---

13In Refs. [146, 69] the combined potential from both the $\Phi_-$ complementary function and particular integral was represented in the form of Eq. (2.30); see Eq. (5) of Ref. [146] and Eq. (5.57) of Ref. [69]. Although this method of presentation was no doubt intended to suppress needless complexity, we believe that the more precise form of (2.42) is helpful. In particular, in (2.42) it is clear that the radial scaling dimension for each term cannot be computed just from knowledge of the representation $L$ for the harmonic with which it appears, as (2.29) and (2.30) would imply. Terms generated from the cross-product between fluxes may occur with a radial scaling dimension $\Delta + \Delta' - 4$ that has no simple relation to the harmonic $E_{LM}$ with which they are partnered in the potential.

14In Agarwal et al. [146], a list of lowest scaling dimensions was given in Eq. (6). However, the value $\sqrt{28} - 3$ in this list should not appear. It is absent from the similar list given in Ref. [141]. We would like to thank Nishant Agarwal for confirmation of this observation.
understanding will require further numerical work that is beyond the scope of this paper.

Summary: the complete brane potential.—To summarize, the total potential for the D3-brane consists of:

- the Coulomb term (2.11),
- the mass term (2.12) generated by coupling to the four-dimensional Ricci scalar,
- the homogeneous terms (2.30) generated by the complementary function for \( \Phi^- \),
- the terms generated by (2.42), with amplitudes modelled as described above, from the particular integral for \( \Phi^- \).

2.2 Experimental procedure

In §2.2.1 we describe our software stack and sampling strategy, and our procedure for collecting observables. In §2.2.2 we explain our definition of an adiabatic limit, based on studying eigenvalues of the mass matrix. Finally, in §2.2.3 we compare the performance of our pipelines and the resulting distributions, and show that (with some caveats) these demonstrate good agreement.

2.2.1 Software stack and sampling strategy

We employ two separate pipelines to harden our analysis against numerical and implementation errors. One pipeline is based on the CppTransport platform [106, 76]. This is a C++ framework for computation of inflationary observables, up to and including those derived from the three-point function, based on an implementation of the ‘transport’ method [108, 109, 79, 75]. The second pipeline is based on PyTransport [80, 77]. This is an independent Python implementation of the same transport system, but making different numerical choices and using a different numerical integrator. Neither pipeline uses the slow-roll approximation, except to set initial conditions for each correlation function.

Although CppTransport and PyTransport are related, they are not equivalent: their implementation details differ, including the exact set of equations that are solved and the underlying computer algebra system used to perform symbolic computations.\(^{15}\) Therefore comparison between these pipelines is not empty. Differences in their output can be regarded as an indication of the ‘implementation error’ from our inability to perform perfectly accurate computations.

At the base of the software stack we use a shared Python script that builds versions of Tables 2.3 and 2.4–2.6 and combines them according to the rules of Eqs. (2.30) and (2.42) to obtain their contribution to the D3-brane potential. We restrict attention to operators with radial scaling dimension \( \delta \leq \Delta_{\text{max}} = 3.8 \), as explained in §§2.1.3–2.1.4 above. The

\(^{15}\)The systems used are GiNaC [179] for CppTransport and SymPy for PyTransport.
script writes out \texttt{CppTransport} and \texttt{PyTransport} model files containing canonical forms for the potential, its first three derivatives, the components of the field-space metric and its inverse, and the components of the field-space Riemann tensor $R^A_{BCD}$. By sharing expressions for these quantities we ensure that both pipelines perform their calculations using the same parametrization, so that subsequent analyses compare like to like. After this stage, symbolic manipulations carried out by the pipelines are independent.

### 2.2.1.1 \texttt{CppTransport} pipeline

The \texttt{CppTransport} translator converts the model file into a custom \texttt{CosmoSIS} module \cite{129}. It expects an input datablock containing the 1,212 parameters of the inflationary model,\footnote{For details of the data flow through a \texttt{CosmoSIS} pipeline, see Ref. \cite{129}.} and uses \texttt{CppTransport}'s internal solver to obtain values for the corresponding background evolution and $n$-point functions. These are written into the outgoing datablock for use by later stages of the pipeline.

This implementation is used to generate our primary trajectory catalogue and compute observables for its members. There are two steps. First, as explained in §2, we generate a catalogue of inflationary solutions by sampling over 450,000,000 trajectories. Our methodology is essentially that proposed by Easther et al. \cite{180}. We use the \texttt{apriori} sampler (part of the default \texttt{CosmoSIS} package) to repeatedly draw realizations of the parameters listed in Table 2.1. Each trajectory is evolved from fixed initial conditions $x = 0.9$ and $\theta_1 = \theta_2 = \phi_1 = \phi_2 = \psi = 1$, with the field velocities set to zero. The observables do not depend significantly on these choices, as we explain in §2.3.3 below. The calculation terminates when either: (1) inflation exits gracefully as $\epsilon$ smoothly approaches unity, or (2) the D3/$\overline{\text{D3}}$ pair dissolve in a hybrid transition, taken to occur when their separation is smaller than $\Delta x = 0.02$. The Lagrangian parameters and Wilson coefficients for the subset of roughly 90,000 trajectories that inflate for $N > 60$ e-folds constitute the required catalogue. A candidate trajectory is rejected if any of the following conditions apply:

- it is not initially inflating,
- it does not reach the hybrid transition that describes D3/$\overline{\text{D3}}$ annihilation while inflation is still ongoing, or within a cutoff of 10,000 e-folds,
- the brane is ejected from the ultraviolet end of the throat,
- the potential becomes negative at any point in the evolution,
- numerical overflow, underflow or an integration error occurs.\footnote{During integration, \texttt{CppTransport} automatically tests for the following error conditions: (a) $H^2$ becomes...}

**Catalogue of observables.**—Second, the completed catalogue is processed to determine

\footnote{An early version of the \texttt{CppTransport} pipeline was written by Sean Butchers, whom we thank for assistance in preparing this section.}
inflationary observables for each trajectory. The calculation is broken into reusable components that are assembled as a second \textit{CosmoSIS} pipeline, controlled programmatically rather than coupled to a sampler. For each entry in the catalogue the pipeline performs the following steps:

**Step 1 — Two-Point Function**

**Power spectra:** It was explained in §2 that the power spectrum in the D3/D3 model is often not scale invariant (or even monotonic), and therefore summary observables measured at a single scale are frequently a poor predictor of the goodness-of-fit to observation. Nevertheless, they have some uses. They allows us to compare with previous analyses, and they are still a convenient way to organize our catalogue of trajectories.

For these reasons we collect summary power spectrum observables $A_s$, $A_t$, and $r$, evaluated at $k_\ast = 0.002\, \text{Mpc}^{-1}$. Here, $A_s \equiv P_\zeta(k_\ast)$ and $A_t \equiv P_h(k_\ast)$ measure (respectively) the amplitude of the dimensionless power spectra for the uniform-density gauge curvature perturbation $\zeta$, and for tensor modes. They are defined in terms of equal-time correlation functions,

$$
\langle \zeta(k)\zeta(k') \rangle = (2\pi)^3 \delta(k + k') P_\zeta(k) = (2\pi)^3 \delta(k + k') \frac{2\pi^2}{k^3} P_\zeta(k),
$$

(2.43a)

$$
\langle h_s(k)h_s'(k') \rangle = (2\pi)^3 \delta(k + k') \delta_{ss'} \frac{\pi^2}{2k^3} P_h(k),
$$

(2.43b)

where $h_s(k)$ is a tensor perturbation of polarization $s$ ($s = +, \times$) in a normalization where the polarization matrices $e^s$ satisfy $\text{tr}(e^s \cdot e^{s'}) = 2\delta_{ss'}$. With this definition the tensor spectrum is conventionally normalized and the tensor-to-scalar ratio satisfies $r \equiv A_t/A_s$.

**Spectral indices:** We evaluate the $\zeta$ spectral index $n_s$ at $k = k_\ast$,

$$
n_s - 1 \equiv \left. \frac{d \ln P_\zeta}{d \ln k} \right|_{k=k_\ast},
$$

(2.44)

Because of the flatness of the tensor spectrum, there are trajectories for which it is not straightforward to collect a reliable numerical estimate of the spectral index $n_t \equiv d \ln P_h/d \ln k$. For a detailed discussion of this and other computational details see §2.2.1.2 below.

**Matching equation:** To relate physical scales to a horizon-exit time we use the matching equation \cite{181, 182, 183},

$$
N(k) = 59.57 - \ln \frac{k}{k_\ast} + \ln \left( \frac{H_k}{10^{16} \text{ GeV}} \frac{M_P^{1/2}}{H_{\text{end}}^{1/2}} \right),
$$

(2.45)

\footnote{See Eq. (20) of Ref. [183], from which we have dropped the slow-roll approximation. Note that there

\begin{itemize}
\item (a) $k$ becoming negative;
\item (b) $\epsilon \equiv -\dot{H}/H^2$ becoming negative;
\item (c) $\epsilon$ becoming greater than 3;
\item (d) $V$ becoming negative;
\item (e) any component of a correlation function becoming $\infty$ or NaN.
\end{itemize}
where \(N(k)\) is the horizon exit time of the physical mode \(k\), measured in e-folds from the end of inflation. The corresponding Hubble rates are \(H_k\) and \(H_{\text{end}}\), respectively. We assume that reheating completes instantaneously, and that decay products from break-up of the scalar condensates thermalize into radiation.

**Step 2 — Three-Point Function**

**Equilateral and folded configurations:** We measure the amplitude of three-point correlations for two indicative \(\langle \zeta \zeta \zeta \rangle\) bispectrum configurations. In the Fergusson–Shellard parametrization,\(^{20}\) these are: (1) an equilateral configuration \(\{k_t = 3k_*, \alpha = 0, \beta = 1/3\}\); and (2) a folded configuration \(\{k_t = 3k_*, \alpha = 0, \beta = 0.005\}\). We report the correlation as a measurement of the reduced bispectrum \(f_{\text{NL}}(k_1, k_2, k_3)\), defined by

\[
f_{\text{NL}}(k_1, k_2, k_3) = \frac{6}{5} \frac{B_\zeta(k_1, k_2, k_3)}{P_\zeta(k_1)P_\zeta(k_2) + \text{cyclic}},
\]

where ‘+ cyclic’ implies that the preceding term is to be summed over cyclic permutations of the momenta \(k_1, k_2, k_3\). The \(\zeta\) bispectrum \(B_\zeta\) satisfies

\[
\langle \zeta(k_1)\zeta(k_2)\zeta(k_3) \rangle = (2\pi)^3 \delta(k_1 + k_2 + k_3)B_\zeta(k_1, k_2, k_3)
\]

and the correlator is computed at equal times. When evaluated on our representative equilateral and folded configurations we denote the reduced bispectrum by \(f_{\text{NL}}^{\text{eq}}\) and \(f_{\text{NL}}^{\text{fold}}\), respectively.

**Squeezed configurations:** As explained in §2 and Appendix C, it is too time-consuming to compute three-point correlations on a squeezed configuration for the entire primary catalogue. Instead, we sample the squeezed configurations \(\{k_t = 3k_*, \alpha = 0, \beta = 0.9\}\) and \(\{k_t = 3k_*, \alpha = 0, \beta = 0.95\}\) on a separate catalogue of trajectories to determine how their amplitudes correlate with the equilateral and folded configurations.

**Step 3 — Adiabatic Limit**

is a minor typo in the version of this equation that appears in Ref. [183]; the correct numerical constant appearing in it should be 55.98, not 55.75. The numerical constant quoted in (2.45) includes this correction. We thank Peter Adshead for helpful correspondence.

\(^{20}\)The momenta \(k_1, k_2, k_3\) that participate in a three-point function such as \(\langle \zeta(k_1)\zeta(k_2)\zeta(k_3) \rangle\) satisfy the ‘triangle’ condition \(k_1 + k_2 + k_3\) as a consequence of statistical translation invariance. This makes the correlator a function only of \(k_1, k_2, k_3\). In the Fergusson & Shellard parametrization we set \(k_t = k_1 + k_2 + k_3\) to be the perimeter of the momentum triangle. Then \([56]\),

\[
\begin{align*}
k_1 &= \frac{k_t}{4}(1 + \alpha + \beta), \\
k_2 &= \frac{k_t}{4}(1 - \alpha + \beta), \\
k_3 &= \frac{k_t}{2}(1 - \beta).
\end{align*}
\]

An equivalent parametrization had earlier been introduced by Rigopoulos, Shellard & van Tent \([55]\).
The mass spectrum is computed from the eigenvalues of the mass matrix \( M^{A B} \) [72, 166, 79],

\[
M^{A B} = \nabla^A \nabla^B V - R^{A}_{B K} \dot{X}^J \dot{X}^K - \frac{3 + \epsilon}{M_P^2} \dot{X}^A \dot{X}^B + \frac{\dot{X}^A \ddot{X}^B + \dot{X}^B \ddot{X}^A}{H M_P^2},
\]

where the \( X^A \) are the scalar fields \( \{ x, \theta_1, \theta_2, \phi_1, \phi_2, \psi \} \) appearing in Eqs. (2.8)–(2.9). Indices on \( \dot{X}^A \) and \( \ddot{X}^A \equiv \dot{X}^B \nabla^B \dot{X}^A \) are raised and lowered using the conifold metric \( G^{AB} \) normalized as in Eq. (2.9), \( \nabla_A \) is the covariant derivative compatible with \( G_{AB} \), and \( R_{ABCD} \) is the Riemann tensor constructed from \( \nabla_A \). We sample these eigenvalues at \( N = 55, N = 2.5, N = 1 \) and \( N = 0 \) e-folds before the end of inflation. In §2.2.2 below we explain how these are used to detect the onset of an adiabatic limit.

**Step 4 — Likelihood Function**

Finally, where possible we compute the CMB likelihood for this trajectory using the Planck2015 likelihood code [184].\(^{21}\) The \( \zeta \) power spectrum is sampled at 100 logarithmically-spaced wavenumbers in the range \( 10^{-6} \text{Mpc}^{-1} \leq k \leq 50 \text{Mpc}^{-1} \). This sample is passed to CLASS via the CosmoSIS pipeline and used to compute the CMB angular power spectra \( C_{\ell}^{TT}, C_{\ell}^{TE} \) and \( C_{\ell}^{EE} \). We do not vary the parameters of the post-inflationary cosmology, which are fixed to their Planck2015 best-fit values [185]. The bundled CosmoSIS planck module is used to calculate the likelihood for all these \( C_{\ell} \). If desired, any other likelihood could be substituted in this step.

**2.2.1.2 Computational issues**

**Ultra slow-roll inflation.**—In the D3/D3 model it is known that extended epochs of inflation are typically realized near an inflexion point in the potential [138, 146, 107]. Therefore we must allow for the possibility that some inflationary trajectories enter a phase of ultra slow-roll dynamics [186, 187, 188, 189, 190]. Such phases are characterized by: (1) \( V' \approx 0 \), (2) a small and rapidly decaying value of \( \epsilon \equiv -\dot{H}/H^2 \), and (3) \( \eta \approx -6 \). Previously, the possibility of an ultra slow-roll phase in this model does not appear to have been considered.

At the level of the background, both CppTransport and PyTransport implement the full scalar field dynamics and therefore capture all ultra slow-roll effects. However, initial conditions for each correlation function are estimated using analytic expressions that assume slow-roll dynamics [75]. This is harmless if the dynamics are close to slow-roll while the momenta characterizing an individual correlation function are exiting the horizon, even if a transition to ultra slow-roll inflation occurs later. If slow-roll does not apply the procedure is still mostly harmless for the two-point function, because the slow-roll result \( P_{\delta \phi} \sim H^2 \) continues

\(^{21}\)In fact, this is not done for all trajectories. In some cases we have \( N \geq 60 \) e-folds of inflation from the initial conditions, but too few total e-folds to allow 4.5 e-folds of subhorizon evolution for the largest scale \( k = 10^{-6} \text{Mpc}^{-1} \) needed for the CLASS computation of the \( C_{\ell} \). In such cases there is a choice between rejecting the trajectory or foregoing the Planck likelihood. We choose the latter.
to apply during ultra slow-roll inflation [186]. Therefore the initial condition will be significantly inaccurate only for modes that exit during a transition between slow-roll and ultra slow-roll phases.

The prospects for the three-point function are less straightforward because the slow-roll initial condition is corrected by terms of order $\eta$ [191, 192, 193], which is large during ultra slow-roll. It follows that numerical three-point functions computed by CppTransport and PyTransport must be treated with caution for scales exiting during an ultra slow-roll phase. Nevertheless, if the calculation starts sufficiently far before horizon exit and the $O(\eta)$ displacements do not take the initial condition out of the basin of attraction of the true solution, we may still expect the results to be valid. We return to this question in §2.3.1.

**Power spectrum amplitudes.**—The amplitudes $A_s$ and $A_t$, and hence $r$, are computed directly. In a transport implementation, numerical accuracy is usually determined by the number of e-folds of subhorizon evolution; see Ref. [75] for a detailed discussion. We use 4.5 e-folds, which (subject to the caveats below for the tensor power spectrum) we find to be a reasonable compromise between accuracy and integration time. We have performed a small number of spot-checks to test convergence with increasing subhorizon e-folds, but these do not show significant improvement: see Fig. 2.2. Accuracy also depends on the choice of stepper. We find that Runge–Kutta (Dormand–Prince and Fehlberg) and Adams–Bashforth–Moulton methods sometimes exhibit instabilities, especially in three-point amplitudes, although in our tests these did not propagate to $\zeta$ observables. The Bulirsch–Stoer variable order method produces fast, high-precision solutions without significant instabilities. However, we do not find that the choice of stepper has a significant effect on our final distributions. For the construction of our primary catalogue we use the Dormand–Prince 4th/5th-order method.

**Numerical precision.**—The integrations needed for the D3/D3 model are complex and involve a large number of intermediate steps. If significant cancellations occur, there is a risk of accumulating inaccuracies from roundoff error. To test whether extended precision is needed we have compared a subsample of 1,000 trajectories from our main catalogue using double and long double precision. In nearly all cases we find only sub-percent shifts in $n_s$. However, for $f_{NL}^{\text{eq}}$ we find $\sim 9\%$ of trajectories exhibit shifts larger than 1%, and $\sim 2\%$ of trajectories exhibit shifts larger than 5%. A handful of trajectories shift by 10% or more. Therefore, while the enhanced precision is not required in most cases, it apparently is needed to accurately predict three-point observables for certain trajectories—but we will see in §2.2.3 that these differences do not seem to be reflected in the distribution over an entire catalogue. Nevertheless, we prefer to use the extended precision calculation out of an abundance of caution.

CppTransport supports arbitrary precision arithmetic (although with a significant performance penalty), so although we have not done so it would be possible to perform the

---

22The meaning of long double is implementation-dependent, but on our production Linux platform
Figure 2.2: Dependence on number $N_{\text{subh}}$ of e-folds of subhorizon evolution for the representative trajectory #32327. Top panel: integration time in seconds. The dotted orange line shows a fit to the exponential dependence $t \sim e^{1.25N}$, which is valid for $N_{\text{subh}} \gtrsim 5.5$. Increasing the number of subhorizon e-folds is very expensive, but the following panels show there are diminishing returns for $N_{\text{subh}} \gtrsim 4.5$. Second panel: Convergence of $P_\zeta$ evaluated at $k = 0.05\, \text{Mpc}^{-1}$. The asymptote is extracted by fitting a function of the form $a - be^{-cN}$ and measuring $a$. The labels show the percentage deviation from this asymptotic value. Third panel: Same as second panel, but for $k^6 B_\zeta(k, k, k)$ measured on an equilateral configuration of side $k = 0.05\, \text{Mpc}^{-1}$. Bottom panel: Same as second panel, but for $f_{\text{NL}}(k, k, k)$ measured on the same equilateral configuration.
calculation with even higher precision than long double to verify that it has properly converged. Similar benefits from use of long double precision are known to occur in models of ultra slow-roll inflation, which has features that are similar to the D3/\overline{D3} model.

Spectral indices.—Computation of spectral indices is more challenging, especially for the tensor power spectrum which is extremely flat. Several numerical strategies are available. When applied to the $\zeta$ spectral index $n_s$ these methods all yield consistent results, but none are entirely satisfactory for the tensor spectral index $n_t$.

To be concrete, we collect the fields $X^A$ and the corresponding momenta $\pi^A \equiv dX^A/dN$ into a single phase-space coordinate $X^A = (X^A, \pi^\beta \equiv dX^B/dN)$ and define the two-point function for $\delta X^A$ to satisfy

$$\langle \delta X^A(k_1) \delta X^B(k_2) \rangle = (2\pi)^3 \delta(k_1 + k_2) \Sigma^{AB}(k),$$

(2.49)

where $k = |k_1| = |k_2|$. The $\zeta$ power spectrum $P_\zeta$ can be written $P_\zeta(k) = N_\alpha N_\beta \Sigma^{AB}(k)$, where explicit expressions for the coefficients $N_\alpha$ are known [72, 108, 70, 79, 75]. They become independent of $k$ on superhorizon scales. The first option is to write a transport equation for the ‘spectral matrix’ $n^{AB} \equiv d\Sigma^{AB}/d\ln k$, which can be used to compute $n_s$ [107, 79, 75],

$$n_s - 1 = 3 + \frac{N_\alpha N_\beta n^{AB}}{N_\gamma N_\delta \Sigma^{CD}}.$$

(2.50)

We briefly review this approach in Appendix B. It is conceptually clean, but as explained in Ref. [79] it can happen that we require more e-folds of subhorizon evolution to obtain good numerical results for $n^{AB}$ than for $\Sigma^{AB}$. This is partly because to compute $n_s - 1$ we effectively subtract the leading term from the right-hand side of (2.50).

A similar expression applies for $n_t$, although the calculation is simpler because no gauge transformation is required. For both $n_s$ and $n_t$ we attempt to accelerate convergence by using initial conditions that include subleading terms in both the slow-roll expansion and $k/(aH)$. We find that Eq. (2.50) gives results for $n_s$ that agree with other methods, but its counterpart for $n_t$ does not always yield good results even with a large number of subhorizon e-folds.

A second option is to fit a function of the form $P = A_n (k/k_*)^n$ to sampled values of the power spectrum $P$ near the pivot scale $k_*$, and extract the spectral index from the fit for $n$. This approach has the advantage that it requires only knowledge of the power spectrum and not the spectral matrix. The disadvantage is that the fit can be thrown off by small inaccuracies in the computed amplitude, perhaps caused by noise or other numerical artefacts. If the spectrum has non-negligible tilt these do not usually affect the measured spectral index. However, for the D3/\overline{D3} model, based on analytic estimates we expect roughly $10^{-14} \lesssim n_t \lesssim 10^{-2}$. Because $n_t$ is so small, fluctuations in excess of $\delta \ln P_h \sim n_t \delta \ln k$ can be present between samples with $k$-spacing $\delta \ln k$ (even with the high-precision Bulirsch–Stoer stepper).

with the GCC compiler this is an 80-bit extended precision format. For comparison, double is a 64-bit format.
Specifically, for some trajectories we find the tensor power spectrum to be contaminated by oscillations of very small amplitude. These can spoil automated measurement of $n_t$. The source of these oscillations is not clear, but they are almost certainly not physical. Their amplitude typically decreases when we allow more subhorizon e-folds. On some trajectories this is sufficient to extinguish the oscillations, but on others their amplitude appears nearly stable. Based on this, we speculate that they are possibly a discretization artefact. If so, the same effect (or a closely related one) may be responsible for the poor outcomes from the $n^{AB}$ transport equation, which include positive values for $n_t$ at some values of $k$. This is incompatible with the strong energy condition, which implies that $H$ should decrease. We see similar results from attempts to fit for $n_t$, which can yield positive values by catching the rising edge of an oscillation. This effect can be mitigated by binning the power spectrum before performing the fit, but it is difficult to do this in an automated way without risking errors from over-smoothing. The significant challenge entailed in obtaining an accurate estimate of $n_t$ was already noted by McAllister et al. [151].

In practice, the sampling pipeline fits a quadratic polynomial to 15 logarithmically-spaced power spectrum samples between $k = 1.986 \times 10^{-3} \text{Mpc}^{-1}$ and $k = 2.014 \times 10^{-3} \text{Mpc}^{-1}$. We have performed spot-checks on roughly 1,200 trajectories to compare the $\zeta$ spectral index computed this way with Eq. (2.50). With 4.5 e-folds of subhorizon evolution we find these are typically consistent within 1%. For a handful of trajectories, we make a further confirmation that these estimates are also consistent with fits performed ‘by hand’ using a different range of $k$. Unfortunately, for $n_t$ we find that these fitting procedures typically disagree. We conclude that computing the tensor spectral index using either method is not acceptable for the D3/D3 model.

A third option is to use an analytic approximation for the tensor power spectrum to estimate $n_t \approx -2\epsilon$, which requires only knowledge of the background. This approach has the advantage that it does not depend on the accuracy with which we can compute the power spectrum amplitude. Conversely, it is blind to information provided by the transport calculation that is not included in the analytic approximation. We find negligible correlation between measurements using our numerical procedure and those obtained from $n_t \approx -2\epsilon$, but this is to be expected in light of the foregoing discussion.

In conclusion, using the estimate $n_t \approx -2\epsilon$ is apparently the least unsatisfactory option. In §2.3 the values of $n_t$ we quote are derived using this method with $\epsilon$ sampled at 60 e-folds before the end of inflation. Based on ‘by hand’ fits to the smoothed tensor power spectrum, we believe the resulting values of $n_t$ are accurate within a factor of 2 (but sometimes much better). For now, it is prudent to treat our $n_t$ estimates with caution.

### 2.2.1.3 PyTransport pipeline

The second pipeline is based on PyTransport. Its main purpose is to perform a number of smaller (~18,000 trajectory) complementary studies that explore the dependence of observables on discrete choices made in §2.1. We also use it to test for consistency with
the principal catalogue. Symbolic computation of the potential and curvature tensors is shared with the CppTransport pipeline, but otherwise there is no code re-use.

There are some important differences. The PyTransport pipeline collects less fine-grained metadata about the trajectory. Observables are mostly computed as explained above, except that to give discrepancies an opportunity to manifest we apply a different fitting prescription for the spectral index. This is based on fitting a spline to five sampled power spectrum values for \( k \)-modes with horizon exit values spaced 0.3 e-folds apart. We use 4.5 e-folds of subhorizon evolution, as above, but PyTransport does not offer an option to change the stepper or value type and therefore we use the built-in Runge–Kutta 4/5th-order solver and double precision arithmetic. PyTransport does not currently implement calculation of the tensor power spectrum, so we do not sample \( r \). The pipeline is not controlled by CosmoSIS, but uses a custom sampling layer that draws parameter combinations from the priors listed in Table 2.1 until a prescribed number of trajectories supporting \( N > 60 \) e-folds of inflation have been sampled. However, the criteria for rejecting trajectories are the same as those given on p. 72. The PyTransport pipeline does not compute \( C_{TT}^\ell \), \( C_{TE}^\ell \), \( C_{EE}^\ell \), or the CMB likelihood function.

Finally, we apply a different prior for \( Q \). Specifically, we choose \( Q \sim \beta'(4.16, 494) \), where the ‘beta-prime’ (or ‘inverted beta’) distribution \( \beta'(\alpha, \beta) \) is characterized by shape parameters \( \alpha, \beta \) and has the probability distribution function \( P(x) = x^{\alpha-1}(1 + x)^{-\alpha-\beta}/B(\alpha, \beta) \), where \( B(x, y) = \Gamma(x)\Gamma(y)/\Gamma(x + y) \) is the Euler \( \beta \)-function. This choice is not motivated by physics, but rather sampling efficiency. We will see below that values of \( Q \) that frequently support \( N > 60 \) e-folds of inflation are tightly clustered. It is this phenomenon that underlies the choice \( Q \sim U(0, 0.04) \) made in Table 2.1, which enhances sampling efficiency but has the drawback that it excludes the region \( Q > 0.04 \) completely. To assist in exploring this region we use the opportunity provided by the PyTransport pipeline to introduce a prior that cannot be implemented using CosmoSIS. The beta-prime distribution samples the region of parameter space that is preferential for obtaining \( N > 60 \) e-folds of inflation, while still exhibiting broad tails that allow less-likely regions to be explored.

We find that the posterior distribution for \( Q \) is completely consistent with the posterior produced from the more restrictive prior used by CppTransport.

In Table 2.7 we list the different ensembles to be analysed using the PyTransport pipeline. In each case the priors match those given in Table 2.1 except for the stated variations. Each sample comprises roughly \( 18 \times 10^3 \) trajectories, except for the \( \alpha = 0 \) study which uses a smaller number of trajectories (\( \sim 5 \times 10^3 \)).

### 2.2.2 The adiabatic limit

To determine whether the observables we collect are related to quantities observable in the CMB or large-scale structure, we must understand whether the dynamics become adiabatic before the end of inflation. Where this happens the perturbations are typically conserved
Table 2.7: ‘Small’ studies performed using the PyTransport pipeline

<table>
<thead>
<tr>
<th>study</th>
<th>prior</th>
</tr>
</thead>
<tbody>
<tr>
<td>initial conditions</td>
<td>( x \sim N(0.9, 0.02); \theta_i \sim U(0, \pi); \phi_i \sim U(0, 2\pi); \psi \sim U(0, 4\pi) )</td>
</tr>
<tr>
<td>size of throat(^a)</td>
<td></td>
</tr>
<tr>
<td>sensitivity to truncation(^b)</td>
<td></td>
</tr>
<tr>
<td>homogeneous potential(^c)</td>
<td></td>
</tr>
<tr>
<td>drop ( V_0 ) uplift(^d)</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) Notice that with our conventions, \( \phi_{UV} \) also appears in the potential via the definition of the dimensionless radial coordinate \( x \); cf. Eqs. (2.11) and (2.13). Therefore variation of \( \phi_{UV} \) does not only vary the size of the throat, but also adjusts the scale of some terms in the potential.

\(^b\) Agarwal et al. studied the potential with the same truncations used here, that is \( \Delta_{\text{max}} \in \{3.8, 3.2\} \) [146]. Dias et al. studied the cases \( \Delta_{\text{max}} \in \{3.8, 3.2\} \) [107].

\(^c\) This prior corresponds to switching off terms in the potential sourced by the flux product \( g_s |\Lambda|^2 / 96 \).

Only the Coulomb term, mass term and the sum of scalar zero-modes (2.30) are retained.

\(^d\) Setting \( \alpha = 0 \) allows us to compare with the prior analyses reported by Agarwal et al. [146], Dias et al. [107], and McAllister et al. [151]. This study is smaller than the others and comprises \( \sim 5 \times 10^3 \) trajectories.

Through the subsequent evolution, provided they remain on superhorizon scales. On the other hand, if the evolution does not become adiabatic then the value of each observable may evolve during and after reheating until all isocurvature modes are exhausted [194, 125, 122, 124, 123, 31, 195, 196]. Here, ‘adiabatic’ has its usual cosmological meaning that there is effectively a single trajectory followed by each patch of spacetime smoothed on some superhorizon scale. The difference between neighbouring patches can only be a time offset \( \delta t \) along this trajectory, from which all other perturbations can be derived.

To diagnose the emergence of an adiabatic trajectory we inspect the eigenvalues of the mass matrix (2.48) collected in Step 3 of §2.2.1.1. Note that the relevant mass matrix is not merely the covariant Hessian \( \nabla A \nabla B V \) that would describe the mass matrix for the scalars alone, but includes mixing with scalar modes of the metric. To obtain the correct mass spectrum it is critical to account for this mixing [197, 198].

On an inflationary trajectory there will typically be one massless or tachyonic mode that is the would-be conserved Goldstone mode \( \zeta \sim \delta \phi_{\text{ad}} / (\sqrt{2} M_P) \) associated with broken time translation invariance along the adiabatic direction [70]. Eigenvectors in the subspace orthogonal to this adiabatic direction span the available isocurvature modes, and their corresponding eigenvalues determine their growth or decay. Fluctuations in a direction with eigenvalue \( m^2 \) typically evolve like \( s(N) = s_0 e^{-\eta(N-N_0)} \), where \( s_0 \) is the amplitude at a fiducial time \( N = N_0 \) and \( \eta = m^2 / (3H^2) \). Therefore fluctuations decay rapidly in any ‘heavy’ direction where the eigenvalue satisfies \( m^2 \gtrsim 3H^2 \).

Exponential suppression implies that isocurvature modes rapidly become small but are never completely extinguished, so there is no unique criterion to determine when a trajectory has become ‘sufficiently’ adiabatic. We choose to sample the mass spectrum at \( N = 0, N = 1 \) and \( N = 2.5 \) e-folds before the end of inflation. The trajectory is de-
declared to be adiabatic if the following conditions apply at all three sample points: (1) one eigenvalue of the mass-matrix is tachyonic, and (2) $N - 1$ eigenvalues are heavy in the sense $m^2/(3H^2) > 1$ \cite{121, 199, 119, 120}.\footnote{This is a sharper criterion than the one proposed in Refs. \cite{107, 120}, in which it was suggested by analogy with the formation of caustics that adiabaticity could be associated with regions where the dilation $\theta$ of a narrowly collimated bundle of trajectories becomes large. For a flow of inflationary trajectories the dilation is approximately given by a normalized sum of eigenvalues $\theta \approx \sum_i m_i^2/H^2$. Therefore the criterion $\theta \gg 1$ is necessary but not sufficient to yield an adiabatic limit in our sense. For example, it can happen that $\theta \gg 1$ but more than one eigenvalue remains light. In this case the trajectories converge onto a sheet rather than degenerating to a single adiabatic trajectory.} This implies a minimum suppression of $e^{-2.5} \approx 8 \times 10^{-2}$ in each isocurvature direction, but usually substantially more.

### 2.2.3 Agreement between pipelines

We now turn to the question of compatibility between the two pipelines, which enables us to assess the integrity of our numerical computations.

**Trajectory-level agreement.**—First, we have performed a number of spot-checks to verify that the pipelines yield compatible results given the same input data. With matching initial conditions and parameter values, we typically find agreement to better than 0.1\% for $n_s$ and better than 0.5\% for the three-point correlation amplitudes. We should regard these as a lower limit on the implementation error for individual trajectories in the catalogue.

**Catalogue-level agreement.**—Second, to test agreement at the level of the catalogue as a whole, we construct a `small’ \texttt{PyTransport} catalogue using the same $\phi_{\text{UV}}, \Delta_{\text{max}}, \text{initial conditions and parameter priors used to construct our primary catalogue. In the left-hand column of Fig. 2.3 we show the resulting distributions of $f_{\text{NL}}^{\text{eq}}$ and $n_s$ for \texttt{CppTransport} (blue) and \texttt{PyTransport} (red). They are qualitatively similar but different in detail. For both observables the most obvious difference is the change in amplitude and location of the peak. In the right-hand column we show the same distributions with the cut $A_s > 10^{-4}$. The amplitude and location of each peak, and the structure of the tails, now show excellent agreement.}

A similar effect can be achieved by cutting out trajectories for which $|f_{\text{NL}}^{\text{eq}}| \lesssim 3 \times 10^{-2}$. This removes the region around the peak of the distribution for $f_{\text{NL}}^{\text{eq}}$, and (although not obvious from Fig. 2.3) the resulting transfer of statistical weight into the tails brings the distributions in agreement. The nontrivial outcome (also for the cut on $A_s$) is that a single cut brings multiple distributions into alignment. The underlying reason, to be demonstrated in §2.3.2.1, is that $A_s, n_s$ and $f_{\text{NL}}^{\text{eq}}$ are all highly correlated in this model. However, choosing to cut on $A_s$ removes marginally fewer trajectories.

The origin of this discrepancy is not completely clear. In the discussion below we enumerate a number of possibilities that we believe are not the cause. In Fig. 2.4 we plot the distribution of $f_{\text{NL}}^{\text{eq}}$ and $n_s$ in the cut region, which clearly exhibits the difference in structure of the peak. The $f_{\text{NL}}^{\text{eq}}$ and $n_s$ distributions are both characterized by a sharp
Figure 2.3: Comparison of distributions for $f_{NL}^{eq}$ and $n_s$ derived from the CppTransport (blue) and PyTransport (red) pipelines. Before cutting on $A_s$, the CppTransport distributions contain 90,039 trajectories and the PyTransport distributions contain 22,453 trajectories. **Left column:** no cuts applied. The distributions are qualitatively similar but disagree in detail. **Right column:** applying the cut $A_s < 10^{-4}$ brings the distributions into agreement. The green hatched region is common to Figs. 2.3, 2.4 and 2.13. Its interpretation is described in the second bullet point on p. 84.
peak and a one-sided tail. The similarity in shape of the distribution is due to the strong correlation between $f_{NL}^{eq}$ and $n_s$. The ‘missing’ tail is so sparsely populated that there are barely any samples beyond the peak, which therefore serves as a cutoff.

We have considered a number of possible explanations for this discrepancy.

- First, it is not caused by disagreement between the pipelines for trajectories that populate the region $f_{NL}^{eq} \gtrsim 0$, $n_s \lesssim 1.0$ where the PyTransport pipeline produces almost no statistical weight. Comparison of output from both pipelines shows excellent agreement for trajectories producing observables in this region.

- Second, one could imagine that small values of $f_{NL}^{eq}$ are simply unreliable because they are dominated by noise. This explanation has the drawback that it would not naturally explain the disagreement in $n_s$. However, it is a possible interpretation for the distribution of $\Delta_{eq \to fold} \equiv f_{NL}^{eq} - f_{NL}^{fold}$ in the right-hand panels of Fig. 2.13 (see below). In a single-field model $|\Delta_{eq \to fold}|$ should be proportional to $n_t$ [57] and therefore negligible on most trajectories, whereas Fig. 2.13 shows that it is typically of order $10^{-2}$. This might happen if each $f_{NL}$ were contaminated by noise at this level. The different behaviour of CppTransport and PyTransport could be ascribed to the differing ODE solvers. The green hatched regions in Figs. 2.3, 2.4 and 2.13 indicate the region that should be excluded in this interpretation. Coincidentally it is roughly the same region that must be excluded to bring the distributions reported by each pipeline into agreement.

However, this interpretation does not seem viable. It is clear from Fig. 2.4 that trajectories for which $f_{NL}^{eq}$ falls within the green hatched region form part of a smooth distribution that extends to much larger values of $|f_{NL}^{eq}|$. In the noise interpretation we would be obliged to assume that this structure can somehow be ascribed to
properties of the noise—and that noise contaminates values of $f_{\text{NL}}^{\text{eq}}$ for which Fig. 2.13 gives no reason to believe it is significant. In our judgement this does not appear probable.

• Third, the discrepant region is associated with large values of $A_s$—indeed, much larger than the observationally-allowed window. At sufficiently large $A_s$ the tree-level approximation will break down, making all predictions unreliable [79, 75]. It is possible (if perhaps unlikely) that $A_s \approx 10^{-4}$ is already large enough for the leading loop correction to become important, especially if the power spectrum runs to large values on small scales.

While this suggests we should already be skeptical of observables computed from trajectories yielding large $A_s$, there seems no reason for CppTransport and PyTransport to fail in different ways if the numerical integration remains under control.

We therefore reject each of these proposed explanations. Instead, we apparently must conclude that the most likely explanation is the performance of the samplers—that is, a systematic difference in the way the pipelines draw parameter combinations leading to trajectories that populate this region. However, we have not managed to identify an error in either pipeline that would cause such a difference. Therefore, to be conservative, we impose the cut $A_s < 10^{-4}$ when discussing observable distributions in §2.3. In this region there is excellent agreement between the pipelines, and we have good reason to believe that the reported distributions are robust.

In any case, as explained above, this cut (or a similar one) is likely to be required to exclude trajectories for which the tree-level approximation is inadequate. Further work is required to improve our understanding of all these issues.

### 2.3 Results

We now present key outputs from our sampling procedure. Agarwal et al. [146] previously discussed the relationship between $n_s$ and $r$, but the reliability of their predictions was unclear in regions of parameter space where the single-field approximation was insufficient. A subsequent analysis performed by Dias et al. [107] yielded comparable results, but also certain differences of detail. Unfortunately, although their computation was more accurate, their catalogue of 564 inflating trajectories was much smaller. McAllister, Renaux-Petel & Xu [151] considered a significantly larger catalogue, but their primary interest was in the frequency of occurrence of multiple-field effects and they did not report distributions for observables. Relationships involving the observables $n_t$, $f_{\text{NL}}^{\text{eq}}$, $f_{\text{NL}}^{\text{fold}}$, $f_{\text{NL}}^\text{eq}$ and $f_{\text{NL}}^{\text{fold}}$ have not yet been studied.

In this section our intentions are twofold. First, we use our primary catalogue of 55,000 trajectories to characterize correlations among the observables $A_s$, $n_s$, $A_t$, $n_t$, $r$, $f_{\text{NL}}^{\text{eq}}$ and $f_{\text{NL}}^{\text{fold}}$. This enables us to compare (up to certain ambiguities) with the results of previous
analyses. However, despite their convenience, we emphasize that these observables often have limited utility. A full likelihood analysis is often needed to determine the goodness-of-fit for each trajectory. Second, the small catalogues listed in Table 2.7 are used to study changes to these distributions when we vary discrete features of the model, such as the truncation of the potential or our choice of initial conditions.

Except for Figs. 2.18–2.17, all distributions reported in this section respect the cut $A_s < 10^{-4}$ discussed in §2.2.3.

### 2.3.1 Background evolution and mass spectrum

**Field evolution.** In Fig. 2.5 we plot the evolution of the background fields for a subset of 100 trajectories as a function of e-folding number $N$ measured from $N = 0$ at the initial data. Typical trajectories show very similar evolution for the radial position $x = r/r_{UV}$, characterized by the onset of rapid motion after a few e-folds followed by an extended loitering period as the inflationary potential flattens at small values of $x$. Very similar behaviour was described by Dias et al. [107]. The angular fields show more variability, but in most cases their values become constant after $\sim 10$ e-folds. This is an indication that trajectories frequently evolve to an adiabatic limit.

To express this quantitatively we apply the criteria for adiabaticity given in §2.2.2, according to which a trajectory is adiabatic if each heavy isocurvature mode satisfies $m^2/H^2 > 3$. Before applying any cuts, we find that 64% of trajectories become adiabatic by the end of inflation. For trajectories that respect the cut $A_s < 10^{-4}$ the corresponding figure is 62%. If the adiabaticity condition is relaxed to $m^2/H^2 > 1$ for the heavy eigenstates, the fraction of adiabatic trajectories increases to $\sim 95\%$. These proportions are consistent between our pipelines and appear roughly consistent with the conclusions of previous studies [146, 107, 151].

**Mass spectrum.** In Fig. 2.6 we show the evolution of the mass spectrum over the period of observable inflation. In each panel we overlay histograms for the ordered eigenvalues of the mass matrix $M^A_B$ given in Eq. (2.48). The top panel shows the mass spectrum 55 e-folds before the end of inflation, which can be regarded roughly as the time of horizon exit for modes contributing to the CMB. The spectrum is relatively closely packed and evenly spaced. The lightest mode is most sharply defined and extends to tachyonic values. The middle and bottom panels show the spectrum at 2.5 e-folds and 1 e-fold before the end of inflation, respectively. In the middle panel, the spectrum is broader and the heavier modes have shifted to slightly higher masses. The lightest mode has become increasingly tachyonic. Similar behaviour was reported in Ref. [151]. Between the middle and bottom panels the distribution of higher-lying heavy modes is stable, but the lightest mode moves even further towards tachyonic values and develops a sizeable gap relative to the rest of the spectrum. This behaviour is expected. On an adiabatic trajectory, $\zeta = \delta \phi / \sqrt{2} \epsilon$ is conserved [125, 31], and $\epsilon$ is typically growing near the end of inflation if there is a
Fig. 2.5: Background evolution for a subsample of 100 realizations. The horizontal scale shows e-folding number $N$ measured from the initial time.

graceful exit. Therefore $\delta \phi$ must also grow, requiring the adiabatic direction in field-space to be a tachyon. Before applying cuts we find that all trajectories exhibit at least one tachyonic mode at 1 e-fold before the end of inflation, but only 0.2% exhibit a second tachyon. There are no trajectories exhibiting three tachyons. The relative occurrence of two tachyonic modes is essentially the same for trajectories that respect the cut $A_s < 10^{-4}$.

Our mass spectra are in qualitative but not quantitative agreement with Dias et al. [107] and McAllister et al. [151]. These references reported mass spectra near horizon exit for a mode contributing to the CMB. Dias et al. did not discuss the spectrum at later times, whereas McAllister et al. found only mild evolution between horizon exit of CMB modes and the end of inflation. In comparison with Dias et al. the shape of our mass distribution at horizon-exit shows good agreement, but the detailed numerical values of the masses are different. This possibly points to a difference in treatment of the ‘particular integral’ modes (2.42) sourced by bulk fluxes, which contribute significantly to the masses of the heavy eigenstates (see §2.3.3).

In comparison with McAllister et al. the numerical values of the masses are similar, but the shape of the distribution is different. We reproduce their conclusion that contributions to the mass matrix from mixing with the metric are generally smaller than contributions from the Hessian $\nabla A \nabla B V$ when CMB scales are leaving the horizon, although the mixing contributions increase in importance towards the end of inflation and are eventually necessary to keep $\zeta$ conserved on an adiabatic trajectory. McAllister et al. observed a mild

\footnote{One might have some reservations regarding the emergence of multiple tachyonic states with large values of $|m^2/H^2|$ given that our pipeline is based on tree-level codes. In single-field inflation, or multiple-field inflation near an adiabatic limit, this is harmless because $\zeta$ is exactly massless and therefore stable (even at loop-level) even though $\delta \phi$ is a tachyon [200, 201]. The situation with multiple tachyons is less clear. In this paper we continue to assume that a tree-level calculation gives an honest representation of the phenomenology, but we note that the issue does not yet appear to have been adequately explored in the literature.}
tachyonic drift of the lightest eigenstate, but at late times the effect is more significant in our realizations. This may be attributable to gravitational mixing. Further, they found that typically the masses $m^2_i$ of the $i = 3, 4$ and $i = 5, 6$ isocurvature states were degenerate at the level of individual realizations. We do not observe this degeneracy, even if we approximate the mass matrix by the Hessian. This apparently points to an underlying difference in the construction of our potentials, perhaps again caused by a differing treatment of the flux-sourced contributions.

**Slow-roll parameters.**—In the upper plot of Fig. 2.7 we plot the distribution of the slow-roll parameters $\epsilon \equiv -\dot{H}/H^2$ and $\eta \equiv d \ln \epsilon / dN$, and their mutual correlation. These parameters are measured at $N = 60$ e-folds prior to the end of inflation, which we denote by the subscript ‘60’. The $\epsilon_{60}$ distribution is bimodal. It would be interesting to understand whether this is related to the effect described by Frazer [61] in which peaks in the distribution function of some observable $O$ are related to critical points in the map $O = O(\theta)$ giving $O$ as a function of the field-space coordinates $\theta$ on a suitable initial hypersurface.

The main weight of the distribution is centred near $\epsilon_{60} \sim 10^{-10}$, which is just a little larger than the typical value $\epsilon \sim 10^{-12}$ reported by McAllister et al. [151], although their evaluation time was not specified. The secondary peak is near $\epsilon_{60} \sim 10^{-4}$. In comparison, McAllister et al. reported only 7% of samples yielded $\epsilon > 10^{-8}$ and no samples with $\epsilon > 10^{-6}$. In our full catalogue we find $\sim 50\%$ of samples yield $\epsilon_{60} > 10^{-8}$ and $\sim 30\%$ yield $\epsilon_{60} > 10^{-6}$. We find no cases where $\epsilon_{60} > 10^{-3}$. The conclusion is apparently that typical values of $\epsilon$ in our catalogue are a few orders of magnitude larger than those reported by Agarwal et al. and McAllister et al. The distribution of $\epsilon$ is also broader.
Figure 2.7: **Top:** Distribution of the slow-roll parameters $\epsilon$ and $\eta$ at 60 e-folds before the end of inflation, and their correlation. **Bottom:** representative time evolution of the $\epsilon$ and $\eta$ parameters. The vertical green dashed line shows the horizon exit time for $k_\star = 0.002\ Mpc^{-1}$. The vertical orange dotted line shows the time at which we sample $\epsilon_{60}, \eta_{60}$. The principal features appearing in the plot are typical, including the slow evolution of $\epsilon$ to very small values before a rapid increase as inflation ends. On this trajectory $|\eta|$ is briefly $\sim 5$ for a period of roughly one e-fold just before horizon exit. In general, although excursions to large positive and negative values are present, they are transient.

This is perhaps related to the inclusion of $\alpha$ in our sampling procedure, which effectively adjusts $H$ while leaving gradients of the potential unchanged.

**Ultra slow-roll inflation.**—Fig. 2.7 demonstrates that, while $\epsilon_{60}$ is always very small, $\eta_{60}$ has excursions to large positive and negative values, although rarely as large as $\eta_{60} \approx 6$. This suggests that full-blown ultra slow-roll is unlikely to occur, although there may be periods during which $\epsilon$ is being suppressed—albeit less dramatically. Inspection of a subsample of trajectories exhibiting large $|\eta_{60}|$ suggests this is the case. In the lower plot of Fig. 2.7 we show the time evolution of $\eta$ for a trajectory belonging to this subsample. Excursions to modestly large $|\eta|$ are present, although on this trajectory they occur before horizon exit of $k_\star$. These excursions are associated with periods during which $\epsilon$ decays in a way similar to the ultra slow-roll phenomenology, but less extreme because $|\eta|$ is not as large.
We have not encountered any trajectories for which the behaviour of $\eta$ clearly supports a diagnosis of full-blown ultra slow-roll. This does not exclude the possibility that, for some trajectories in our catalogue, the initial conditions for observables might be affected by transiently large $|\eta|$. A full analysis of these effects, if they occur, is beyond the scope of this paper. Here we only note that in both the full catalogue and the subsample satisfying WMAP7 constraints on $A_s$ at $3\sigma$, no more than $\sim 5\%$ of trajectories exhibit $|\eta_{60}| > 2$. Assuming this fraction is representative of the proportion of trajectories that could be affected, ultra slow-roll-like effects appear unlikely to distort the final distribution of observables.

$Q$ parameter and Wilson coefficients.—We now consider the posterior distribution of $Q$ and $\alpha$, after applying the cut $A_s < 10^{-4}$. (For definitions, see the bottom two lines of Table 2.1.) Both parameters affect the relative scale of terms in the potential, and therefore influence the likelihood of finding a ‘delicate’ region of field space where the potential is sufficiently flat to inflate.

The left panel of Fig. 2.8 shows that the probability of obtaining an extended epoch of inflation depends strongly on $Q$, with successful realizations clustering tightly around $Q \sim 0.006$. This differs from the value $Q \sim 0.04$ reported by Agarwal et al. [146]. However, as explained in that reference, the narrow range of $Q$ for which prolonged inflation can be realized reflects the need to carefully balance Coulomb attraction with repulsion from the bulk contributions for typical values of the Wilson coefficients $C_{LM}$. Repulsion from the bulk terms scales with $Q$, and the precise point of balance depends on how many terms are retained. The numerical value of $Q$ therefore has no physical significance. However, the discrepancy supports our suggestion of a systematic difference between typical trajectories in our catalogue and those of Refs. [146, 107]. It is not yet clear whether the difference in $Q$ is caused by the same difference responsible for the difference in typical values of $\epsilon$.

Meanwhile, the $\alpha$ distribution is very roughly flat on both sides of $\alpha = 0$, with a small bias.
2.3.2 Two- and three-point observables

We are now in a position to examine the correlation between the summary statistics $A_s$, $A_t$, $n_s$, $n_t$, $r$, $f_{NL}^s$, $f_{NL}^{old}$, and $f_{NL}^{sq}$. To be clear, we recall that these are defined in Eqs. (2.43a)–(2.43b), (2.44), and (2.46). Only the relationship between the scalar spectral index $n_s$ and the tensor-to-scalar ratio $r$ has previously been studied [146, 107].

2.3.2.1 Two-point observables

Failure of scale invariance and monotonicity.—In Fig. 2.9 we plot the relationship between $n_s$ and $r$ for the subsample in which $A_s$ is compatible with the 3σ WMAP7 constraint.

---

Although this is not said explicitly in either reference, we understand it to be the case. We thank
\[ A_s = (2.43 \pm 0.33) \times 10^{-9} \] [149]. This distribution enables a comparison with previous analyses.

We highlight the WMAP7 $3\sigma$ region \( n_s = 0.963 \pm 0.014 \) in red [149, 150]. It is populated only by a handful of trajectories, subject to the caveats mentioned below. This is somewhat surprising. In our other catalogues, the proportion of trajectories that fall within the WMAP7-allowed region is relatively larger, perhaps by a factor of $O(10)$. Presumably, this happens because the allowed region corresponds to such a small fraction of the model’s large parameter space that our sampling is not entirely representative even with 55,000 trajectories. However, in this paper our aim is not to optimize the fit to current datasets, but to explore the statistical distribution of observables for typical values of the parameters appearing in the Lagrangian. We expect the sampling in these typical regions to be more accurately representative.

In Fig. 2.9 the trajectories that fall in the allowed region are highlighted by enclosing cyan squares. One of these is adiabatic; the other two are non-adiabatic. The adiabatic trajectory has an unusually large amplitude of three-point correlations, \( f_{\text{NL}}^{\text{ad}} = -0.749 \), to be discussed in §2.3.2.2 below. One might expect the best-fit trajectory to be one of these three. However, according to the Planck2015 $TT+TE+EE$ likelihood, the best-fit is Trajectory #43,942 with log-likelihood \( \ln L \sim -600 \). It produces a spectrum for which \( n_s \) measured at \( k = 0.002 \text{Mpc}^{-1} \) is marginally blue and does not fall in the WMAP7 $3\sigma$-region at all. This trajectory is highlighted by an enclosing dark blue circle. Note that the general trend with \( \alpha \) is opposite to the catalogue as a whole, in which increasing \( \alpha \) is correlated with decreasing \( \epsilon \sim r/8 \) as explained above. In Fig. 2.9 the smallest \( r \) are correlated with the smallest \( \alpha \).

In Fig. 2.10 we plot the primordial power spectrum for the best-fit trajectory, together with the corresponding angular spectra \( C^T_T \), \( C^E_T \), \( C^E_E \), and the Planck2015 allowed region at $1\sigma$ [184]. In Fig. 2.11 we plot the same quantities for Trajectory #88,167, which is one of the non-adiabatic trajectories that falls in the WMAP7 $3\sigma$ region for \( A_s \) and \( n_s \). For ease of comparison these trajectories are labelled in Figs. 2.9, 2.12 and 2.13.

The situation is clear. Fig. 2.11 demonstrates that Trajectory #88,167 produces a scale dependent, non-monotonic power spectrum. Its form is similar to a portion of the characteristic non-monotonic shape emphasized by Dias et al. [107]. Strictly, the angular spectra \( C^{XY}_\ell \) computed from this primordial spectrum are unreliable, because \( P_\zeta(k) \) will be modified by quenching of isocurvature modes before it is communicated to the CMB. However, for the present discussion this is not of primary concern. What is important is that the apparent near scale invariance suggested by measurement of \( n_s \) at \( k = 0.002 \text{Mpc}^{-1} \) is evidently fictitious; in fact, \( n_s \) varies significantly over the observable range. Our estimates for the \( C_\ell \) show that this trajectory significantly overpredicts the correlation amplitude for $TT$ and $EE$.

Conversely, the best-fit trajectory #43,942 produces a nearly featureless power-law spec-

Nishant Agarwal for helpful correspondence on this issue.
Figure 2.10: Power spectra for our best fit trajectory #43,942. **Left:** angular spectra $C_{\ell}^{TT}$, $C_{\ell}^{TE}$ and $C_{\ell}^{EE}$ for the temperature and $E$-mode fluctuations, and their cross-correlation. **Right:** primordial dimensionless power spectrum $P_\zeta(k) = k^3P(k)/(2\pi^2)$. The power-law fit at $k = 0.002$ Mpc$^{-1}$ is marked. Note that the primordial spectrum has very little running, and is accurately given by the power-law fit over a wide range $10^{-6}$ Mpc$^{-1} \lesssim k \lesssim 50$ Mpc$^{-1}$. The green shaded region highlights the scales $0.002$ Mpc$^{-1} \lesssim k \lesssim 0.2$ Mpc$^{-1}$ that (approximately) contribute significantly to the angular power spectrum for $\ell \lesssim 2000$. 
Figure 2.11: Power spectra for Trajectory #88, 167. The left and right panels match Fig. 2.10.
trum over the entire observable range of $k$; only a very small running is visible for $k \lesssim 10^{-4}\text{Mpc}^{-1}$. From the inset zoom panel in $TT$ it can be seen that although the overall fit is good, the trajectory very slightly overpredicts the amplitude near the third peak, but (not shown) underpredicts near the first peak—as might be expected for a blue primordial spectrum. It is probable that further trajectories can be found that yield an even better fit. Dias et al. already observed that nonmonotonic $P_\zeta(k)$ occur relatively frequently in this model due to the characteristic behaviour of $\epsilon$ as trajectories roll towards, through and away from the inflexion point. We discuss this in more detail on p. 96 below. To correctly assess the goodness-of-fit for these examples we cannot rely on summary statistics such as $A_s$ and $n_s$, but instead require a realistic likelihood calculation.

Note that because $H$ is usually close to constant while observable scales are leaving the horizon, the tensor power spectrum is commonly featureless. Therefore $r$ will inherit scale dependence from $P_\zeta(k)$, and—if it were not so small—predictions for the observability of gravitational waves would also require careful treatment.

The $n_s-r$ correlation.—Fig. 2.9 should be compared with Fig. 9 of Agarwal et al. [146] and Fig. 17 of Dias et al. (arXiv version 2) [107]. The same qualitative features are visible in all these plots, but there are quantitative differences. In Fig. 2.9 the general trend is for larger values of $n_s$ to be correlated with larger values of $r$, with no clear separation between adiabatic and non-adiabatic trajectories. The samples fill out a wedge-shaped region, producing considerable scatter for very red values of $n_s$. Most trajectories cluster in the opposite limit near $n_s \approx 2$, where the spectrum is very blue. These trajectories have an unfavourable CMB likelihood.

In Agarwal et al. the observed values of $n_s$ cover $0.93 \lesssim n_s \lesssim 1.10$ and the corresponding values of $r$ cover $-14 \lesssim \log_{10} r \lesssim -11$. Our values of $n_s$ are typically rather more blue, which could be attributed to the single-field approximation used by Agarwal et al. In Dias et al. [107] the corresponding range of $r$ is not clear because of the choice of axes, but is plausibly $-9 \lesssim \log_{10} r \lesssim -7$; the authors of Ref. [107] did not comment on the discrepancy with Ref. [146]. Our range of $r$ is different again but closer to Dias et al., spanning roughly $-4.6 \lesssim \log_{10} r \lesssim -7.0$. The reason for this substantial variation in $r$ between different analyses is not clear. Trajectories with small $\epsilon$ in the range $-12 \lesssim \log_{10} \epsilon \lesssim -10$ are present in our catalogue, which would correspond to values of $\log_{10} r$ much closer to those of Agarwal et al. However, they are associated with $A_s$ that are outside the observable window. The $\alpha = 0$ study discussed in §2.3.3 suggests that the distribution of observables is not significantly affected by our inclusion of $\alpha \neq 0$ during sampling, so the discrepancy is apparently caused a structural difference in the potential or a difference in sampling methodology.

One possible explanation is variation of the energy scale at which the brane ‘loiter’ near the inflexion point in its potential. The inflexion point is itself a consequence of delicate cancellations between the attractive Coulomb force and repulsive bulk contributions, and (as has already been said) the exact location of the balance point depends on the treatment
of the bulk terms. Small changes to the cancellations that produce the inflexion point could perhaps change the corresponding value of $V$ significantly.

Second, we sample parameters at the fixed scale $k_\star = 0.002 \text{Mpc}^{-1}$, yielding a range of horizon-exit times corresponding roughly to $55 \lesssim N_\star \lesssim 57$, where $N_\star$ measures horizon exit of the scale $k_\star$ in terms of e-folds prior to the end of inflation. Agarwal et al. reported values at the fixed horizon exit time $N_\star = 60$ [146], and Dias et al. used $N_\star = 55$ [107]. For slow-roll inflation such small shifts in the evaluation time often lead only to small shifts in observables, but in this model the character of the trajectories can change depending on the relative position of the initial conditions and the inflexion point. (For example, compare the power spectrum for Trajectory #88, 167 given in Fig. 2.11.) A systematic difference in the evaluation point could perhaps modify statistical properties of the observables.

Finally, Agarwal et al. used a simple single-field approximation to compute $A_s$ which will produce unreliable estimates where multiple-field effects are significant. We have already noted that the same applies to the spectral index, especially if the power spectrum is not close to scale invariance. This may explain the different range of $n_s$ observed between our catalogues. To decide which of these possible causes, if any, contribute significantly to the differences in the $n_s$–$r$ relation would require a forensic analysis of each implementation, which is beyond the scope of this paper. We comment further on these issues in §2.4.

Moving beyond the difference in normalization, the form of the $n_s$–$r$ relationship is qualitatively similar in all analyses. In Agarwal et al. the correlation is tightest and the relationship is approximately linear. In Dias et al. and our Fig. 2.9 the correlation is compatible with linearity, but there is considerable scatter and it is not clear that other functional forms are excluded. Assuming linearity, however, the slope of the relationship differs between analyses with our Fig. 2.9 being shallowest and Dias et al. being steepest.

Red and blue spectral indices.—The majority of our trajectories yield blue spectral indices, although Figs. 2.10 and 2.11 show that any numerical value for $n_s$ should be interpreted with care. The same effect was reported by Agarwal et al. [146], Dias et al. [202] and McAllister et al. [151], who all observed that when $\epsilon \ll 1$ (which is the case in the vicinity of the inflexion point), the spectral index inherits its sign from $V''$, where $'$ denotes a derivative in the adiabatic direction in field space. Therefore, modes leaving the horizon before the inflexion point are blue-tilted, whereas those leaving the horizon after the inflexion point are red-tilted. This is one cause of non-monotonicity in the power spectrum.

Agarwal et al., Dias et al. and McAllister et al. noted that this effect limited the number of trajectories producing red $n_s$, because it is more difficult to obtain sufficient e-folds of inflation after crossing the inflexion point. Our sample contains populations of trajectories with red and blue values of $n_s$ that conform to these expectations, which is a primary driver for the preponderance of blue values $n_s > 1$ in Fig. 2.9. We also find a relatively larger population of trajectories yielding red $n_s$ where horizon exit occurs prior to crossing the
Figure 2.12: Tensor-to-scalar ratio \( r \) against \( n_t \) for trajectories satisfying the WMAP7 constraint on \( A_s \) at 3\( \sigma \). The orange dotted line shows the consistency ‘bound’ \( r = 8|n_t| \), and the points are colour-coded by their value of \( \alpha \). The best-fit trajectory is highlighted with a large open circle, as in Fig. 2.9.

inflexion point, but before approaching an adiabatic limit. This invalidates expectations based on the sign of \( V'' \). Further, these trajectories typically pass through a sequence of critical points where slow-roll may not be a good approximation. This population of red-\( n_s \) trajectories does not appear to have been identified in previous analyses. However, their spectra clearly cannot be monotonic, so it is not yet clear whether their detailed properties can be observationally acceptable.

**Consistency equation.**—In single field models the tensor-to-scalar ratio and tensor spectral index are related to leading order in slow-roll by the ‘consistency relation’ \( r = -8n_t \) [203, 204]. In multiple-field models this is weakened to an inequality \( r \leq -8n_t \), also valid only to leading order in slow-roll and assuming that all modes contributing to \( \zeta \) are massless\(^{27} \) [72]. This follows from the Cauchy–Schwarz inequality applied to the projection \( \zeta = N_A \delta \chi^A \) from field fluctuations \( \delta \chi^A = (\delta X^A, \delta \pi^B) \) onto \( \zeta \). In Fig. 2.12 we plot \( r \) against our estimated \( n_t \) for trajectories satisfying the WMAP7 bound on \( A_s \) at 3\( \sigma \). The ‘consistency bound’ is represented by the orange dotted line, and is respected by a clear majority of trajectories. A small number of trajectories exhibit marginal transgressions. The most likely explanation is that these are effectively single-field models that should lie exactly on the bound, but our procedure for estimating \( n_t \) has produced a result that is slightly too small.

In the full catalogue we observe a population of trajectories that exhibit more significant (but still not dramatic) transgression of the ‘consistency bound’. There are several possible explanations of this effect, including misprediction of \( n_t \) as suggested above. However, it is also possible that these are trajectories for which the massless approximation fails and

\(^{27}\)The massless condition was not discussed in Ref. [72] and has frequently not been stated in the literature, but it is required. In the analysis of Ref. [72] it appears in the assumption that the scalar two-point function is proportional to the kinetic mixing matrix. This need not be true if the fields have a non-negligible mass matrix, as may be the case in the D3/D3 model. For a similar discussion from a different perspective, see §4.3.4 of Ref. [151].
the fields become anticorrelated at horizon exit, leading to a reduction in the final power spectrum amplitude below what would be predicted based on the adiabatic mode alone. If this reduction is sufficiently dramatic it could cause a violation of the massless consistency bound.

This anticorrelation was observed by McAllister et al., who described it as ‘destructive interference’ [151]; see the discussion in §4.3.4 of this reference. Our results apparently reproduce their observations. After performing ‘by hand’ fits to the tensor power spectrum in order to obtain the most accurate possible measurements of \( n_t \), we find that residual violations of the consistency bound are no more than a few percent. It is not yet clear whether this is a genuine effect caused by violation of the massless condition, or is simply an unmodelled error in fitting to the tensor power spectrum. To produce a convincing demonstration of violation of the consistency bound, or otherwise, would require an investment in higher-quality numerical computations of the tensor spectral index. It would be interesting to return to this question in future work.

### 2.3.2.2 Three-point correlations

**Equilateral and folded configurations.**—Next consider the amplitude of three-point correlations, measured by \( f_{NL}^{eq} \) and \( f_{NL}^{fold} \), which represent the major new results presented in this paper. (Recall that we sample these quantities at a fixed scale \( k_t = 3k_x = 3 \times 0.002 \text{ Mpc}^{-1} \). For details of our observables, see §2.2.1.1.) In general \( f_{NL}^{eq} \) and \( f_{NL}^{fold} \) need have no simple relation, but in a single-field model they are connected by Maldacena’s formula\(^{28}\)

\[
  f_{NL}(k_1, k_2, k_3) = -\frac{5}{12(n_s - 1)} + g(k_1, k_2, k_3)n_t, \tag{2.51}
\]

where \( g(k_1, k_2, k_3) \) is a calculable function of the momenta \( k_i \) [57], but does not depend on slow-roll parameters or other model-dependent data. Hence, in a single-field model, the ‘residual’ \( \Delta(\text{eq} \rightarrow \text{fold}) \equiv f_{NL}^{eq} - f_{NL}^{fold} \) is proportional to \( n_t \) multiplied by a fixed factor depending only on the momentum configuration.

In Fig. 2.13 we plot \( f_{NL}^{eq} \) and \( f_{NL}^{fold} \) for trajectories that satisfy the WMAP7 constraint on \( A_s \) within 3\( \sigma \). As in Figs. 2.9 and 2.12 the points are colour-coded by their value of \( \alpha \), with values near \( \alpha = +1 \) coded green and values near \( \alpha = -1 \) coded purple. Previous analyses have assumed \( \alpha = 0 \). We highlight the Planck2015 \( TT+TE+EE \) best-fit trajectory, described above, which has small amplitudes \( f_{NL}^{eq} \approx -4.73 \times 10^{-3} \) and \( f_{NL}^{fold} \approx -4.75 \times 10^{-3} \). We will see below that these values are typical when both \( A_s \) and \( n_s \) fall in the observationally-allowed window.

The top-left panel of Fig. 2.13 shows tight correlation between \( f_{NL}^{fold} \) and \( f_{NL}^{eq} \), as would be predicted by (2.51). In single-field cases the small scatter is due to the smallness of \( n_t \). However, even in cases where multiple-field effects are important, this panel shows that

\[^{28}\text{Compare Eq. (4.9) of Maldacena, arXiv version 5 [57]. Note there is a sign flip of } f_{NL} \text{ between this equation and Eq. (2.51), because Ref. [57] defined } f_{NL} \text{ with the opposite sign convention.}\]
Figure 2.13: Relation between three-point correlation amplitudes in equilateral and folded configurations with $k_0 = 3 k_\star = 3 \times 0.002 \text{ Mpc}^{-1}$. (For details of the three-point configurations we sample, see §2.2.1.1.) Plotted points correspond to trajectories that satisfy the 3σ WMAP7 constraint on $A_s$. Except in the top-left panel, the populations with $\alpha > 0$ and $\alpha < 0$ are plotted separately to aid comparison. **Top left:** $f_{\text{NL}}^{\text{fold}}$ against $f_{\text{NL}}^{\text{eq}}$. The dashed orange line has the functional form $f_{\text{NL}}^{\text{fold}} = a + f_{\text{NL}}^{\text{eq}}$, where $a = 8.3 \times 10^{-3}$ is fit to the measured values with correlation coefficient $R > 0.995$. **Bottom left:** $f_{\text{NL}}^{\text{eq}}$ against $n_s$. The dashed orange line has functional form $f_{\text{NL}}^{\text{eq}} = b - (5/12)(n_s - 1)$ as in Eq. (2.51). For $\alpha > 0$ we find $b = -0.0270$ and for $\alpha < 0$ we find $b = -0.00286$. **Top right:** the ‘residual’ $|\Delta(\text{eq} \rightarrow \text{fold})| \equiv |f_{\text{NL}}^{\text{eq}} - f_{\text{NL}}^{\text{fold}}|$ against $n_t$ (estimated from $\epsilon_{60}$ as described in §2.2.1.2). **Bottom right:** $|\Delta(\text{eq} \rightarrow \text{fold})|$ against $N_\star$, where $N_\star$ measures the horizon exit time (in e-folds before the end of inflation) of the pivot scale $k_\star = 0.002 \text{ Mpc}^{-1}$. The interpretation of the green hatched region was discussed in §2.2.3. **Colour bar:** in each panel, points are colour-coded by their value of $\alpha$. In all four panels the best-fit trajectory is highlighted as in Figs. 2.9–2.12.
analytic approximation of \( n \) function of our estimated \( n_s \). In the top-right panel we plot the absolute value of the ‘residual’ (bottom plot) separately. In both cases a group of effectively single-field trajectories is visible that accurately match the dashed lines, each of which is chosen to have the slope \(-5/12\) predicted by (2.51). This is especially clear for small values of \( n_s \), where almost all points cluster close to the line. For larger values of \( n_s \) the scatter becomes more significant, especially for \( \alpha > 0 \), and may indicate that multiple-field effects are relevant in this region. Even here, however, the \(-5/12\) dependence predicted by (2.51) is strikingly well reproduced. Clearly, this dependence is the principal driver for large values of \( |f_{\text{NL}}| \). A notable exception is the single adiabatic trajectory in the WMAP7 3σ regions for both \( A_s \) and \( n_s \) (see Fig. 2.9). This has \( n_s - 1 \approx -0.06 \) but a large amplitude \( f_{\text{NL}}^{eq} = -0.749 \), derived from an abrupt change of angular minimum to be discussed in §2.3.2.4. However, the plot shows that examples of this kind are relatively rare.

In the top-right panel we plot the absolute value of the ‘residual’ \( |\Delta(\text{eq} \rightarrow \text{fold})| \) as a function of our estimated \( n_t \). To be clear, we repeat that these are computed using the analytic approximation \( n_t = -2\epsilon_{g0} \). According to Eq. (2.51), \( |\Delta(\text{eq} \rightarrow \text{fold})| \) should be proportional to \( n_t \), which on logarithmic axes would correspond to a line with unit slope. This is not what we find; instead, \( |\Delta(\text{eq} \rightarrow \text{fold})| \approx 10^{-2} \) is roughly constant, independent of \( n_t \). Here there is a clear segmentation of the populations \( \alpha \gtrsim 0 \). The \( \alpha > 0 \) population has more scatter and is restricted to \( n_t \gtrsim 10^{-6} \). The \( \alpha < 0 \) population extends (with decreasing density) to \( n_t \sim 10^{-8} \) with roughly constant amplitude.

The possibility that this behaviour is caused by contamination from numerical noise was rejected in §2.2.3. Instead, we must apparently attribute it to multiple-field effects. The size of the effect is comparable to the running from equilateral to squeezed configurations to be discussed in §2.3.2.3 below, which is consistent with a multiple-field origin.

Finally, the bottom-right panel shows \( |\Delta(\text{eq} \rightarrow \text{fold})| \) as a function of \( N_* \), the horizon-exit time for \( k_* = 0.002 \text{Mpc}^{-1} \). The structure is very similar to the top-right panel, but the segmentation is even clearer with each population confined to nearly exclusive regions. The \( \alpha > 0 \) population is restricted to \( 56.6 \lesssim N_* \lesssim 56.8 \), whereas the \( \alpha < 0 \) population is restricted to the wider range \( 55.3 \lesssim N_* \lesssim 56.6 \).

Amplitude distribution for \( |f_{\text{NL}}| \).—A key question is the typical amplitude of \( f_{\text{NL}}(k_1, k_2, k_3) \) on observationally accessible configurations, because this determines whether non-Gaussian effects are detectable. In Fig. 2.14 we plot distribution functions for \( |f_{\text{NL}}^{eq}| \). (We do not give separate distributions for \( f_{\text{NL}}^{\text{fold}} \) because Fig. 2.13 shows it to be highly correlated with
Distribution functions for $|f_{\text{NL}}^\text{eq}|$. Left: probability density function $P\left(\frac{|f_{\text{NL}}^\text{eq}|}{dx} \in (x, x + dx)\right)$ for different cuts on the catalogue. The blue histogram applies the minimal cut $A_s < 10^{-4}$. The red histogram represents the subsample of trajectories satisfying the WMAP7 constraint on $A_s$ at 3$\sigma$. Right: Tail distribution $P\left(|f_{\text{NL}}^\text{eq}| > x\right)$, measured using different cuts on $A_s$. The orange horizontal dotted line marks the point where $P\left(|f_{\text{NL}}^\text{eq}| > x\right) = 0.45$, which for the WMAP7 cut corresponds roughly to $|f_{\text{NL}}^\text{eq}| > 0.5$.

$f_{\text{NL}}^\text{eq}$

Fig. 2.9 shows that many trajectories yield $n_s \approx 2$, so the estimate $f_{\text{NL}} \approx -(5/12)(n_s - 1)$ suggests we should expect a concentration near $f_{\text{NL}} \approx -0.8$. It should be borne in mind that bispectra satisfying this estimate will be strongly scale dependent whenever the spectrum, and likewise $n_s$, are strongly scale dependent.

This expectation is approximately borne out the detailed distribution for $|f_{\text{NL}}^\text{eq}|$ given in the left panel of Fig. 2.14. For the minimal cut $A_s < 10^{-4}$ (blue), the distribution is rather flat for $0.1 \lesssim |f_{\text{NL}}^\text{eq}| \lesssim 0.6$. There is a gently decaying tail to smaller absolute values, and more abrupt decay for larger values. There is almost no weight in the distribution for $|f_{\text{NL}}^\text{eq}| > 0.8$, as can be seen by comparison with the right panel showing the tail distribution.

For the subsample of trajectories that satisfy the WMAP7 constraint on $A_s$ at 3$\sigma$ (red), the distribution is clustered in the region $0.5 \lesssim |f_{\text{NL}}^\text{eq}| \lesssim 0.6$, as suggested by the distribution for $n_s$. The tail to low absolute values $|f_{\text{NL}}^\text{eq}| \lesssim 0.1$ is heavily depopulated and there are only a handful of samples for which $|f_{\text{NL}}^\text{eq}| \gtrsim 1$. In the right-hand panel, the orange horizontal dotted line marks the point where $P\left(|f_{\text{NL}}^\text{eq}| > x\right) = 0.45$, chosen because for the WMAP7 sample it corresponds roughly to $|f_{\text{NL}}^\text{eq}| > 0.5$. In the full catalogue $|f_{\text{NL}}^\text{eq}|$ exceeds 0.5 much less frequently, in no more than 8% of cases. This reflects an approximate trend within our catalogue (already mentioned in §2.2.3), that larger values of $A_s$ are correlated with smaller values of $f_{\text{NL}}$.

Unfortunately these large $|f_{\text{NL}}|$ can not be regarded as an observable signature in our own universe, because they derive from values of $n_s - 1$ of order unity. For trajectories that satisfy observable constraints on both $A_s$ and $n_s$, it appears that $|f_{\text{NL}}|$ will typically be small—as is the case for our best-fit trajectory.
2.3.2.3 Squeezed configurations

Finally, we turn to squeezed configurations. In Appendix C we show that modestly squeezed isosceles configurations with \( \beta = 0.9 \) and \( \beta = 0.95 \) (corresponding to squeezings \( k_3/k_t = 0.05 \) and 0.025, respectively) require integration times in the range 1,000 s to 2,000 s. Taking the CMB to receive contributions from approximately 0.005 Mpc\(^{-1}\) to 0.2 Mpc\(^{-1}\), the maximum observable squeezing is roughly \( k_3/k_t \approx 0.0125 \) or \( \beta \approx 0.975 \). Therefore these estimates can be taken as a reasonably reliable lower limit on the compute time required for predictions in the observational range. This would increase processing time for the entire catalogue by a factor perhaps in the range 3 to 5, which is prohibitive given the already sizeable computational demands for obtaining observables.

Gelaton and QSFI effects.—Instead, we study the relationship between \( f^\text{NL} \) and \( \{f^\text{sq}, f^\text{eq}\} \) on a smaller sample of roughly 3,000 trajectories. In the top-left panel of Fig. 2.15 we plot the computed value of \( f^\text{eq} \) against \( f^\text{eq} \) for trajectories that satisfy the WMAP7 bound on \( A_s \) at 3\( \sigma \). As expected this shows strong correlation, which applies equally to the full catalogue satisfying \( A_s < 10^{-4} \). The same conclusion applies to \( f^\text{fold} \), which itself correlates strongly with \( f^\text{NL} \). The orange dashed line corresponds to the approximate relation \( f^\text{eq} = -0.006 + 0.92 f^\text{eq} \), from which we conclude that the typical amplitude of \( f^\text{eq} \) is very similar to \( f^\text{eq} \) but just a little smaller. The characteristic clustering of values between \(-0.6 < f^\text{eq} < -0.1 \) is clearly visible (cf. the blue distribution in Fig. 2.14).

In the top right panel of Fig. 2.15 we plot the shift \( \Delta (\text{sq} \rightarrow \text{eq}) = f^\text{eq} - f^\text{NL} \) between squeezed and equilateral configurations, expressed as a percentage of the equilateral amplitude. The shift almost always falls between \(-4\% \) and \(-6\% \), with values near \(-4\% \) being favoured. In absolute values this corresponds to \( \Delta (\text{sq} \rightarrow \text{eq}) \approx \text{few} \times 10^{-3} \). In models for which the slow-roll approximation applies throughout, we expect the slow-roll parameters to set the scale of the shift [57, 206]. For example, in the axion–quadratic model studied in Ref. [75] the shift computed between the same configurations is roughly \( \sim \text{few} \times 10^{-2} \), which is comparable to \( \eta = d \ln \epsilon / dN \). From this point of view the typical value of \( \Delta \) observed in the D3/D3 model is not unusually large, and indeed is comparable to \( \eta \) in many realizations (see Fig. 2.7 and also the discussion of the residual \( \Delta (\text{eq} \rightarrow \text{fold}) \) above).

It is possible that some of the more significant shifts of 10\% or larger are associated with an unusual change in amplitude near equilateral configurations, which might be expected from a ‘gelaton’-like or ‘QSFI’-like scenario [207, 152, 153, 154]. Assuming the rate 0.07\% suggested by McAllister et al., we would expect to observe perhaps O(10) QSFI-like examples in a catalogue of this size [151]. It would be exceptionally interesting (but numerically expensive) to compute exact bispectrum shapes for the trajectories that exhibit the largest shifts between equilateral and squeezed configurations.

Shape dependence.—In general the bispectrum may depend on shape and scale through the ratios \( k_i/k_* \) and \( k_i/k_t \) for \( i = 1, 2, 3 \). As we move from equilateral to squeezed configurations, taking \( k_3 \) to be the squeezed momentum, the ‘shape’ ratios \( k_1/k_t, k_2/k_t \) vary
Non-Gaussianity in D3-brane inflation

Figure 2.15: Analysis of the squeezed amplitude $f_{NL}^{sq}$. We impose our standard cut $A_s < 10^{-4}$. **Top left**: $f_{NL}^{sq}$ against $f_{NL}^{eq}$. Green points correspond to $\beta = 0.9$ and squeezing $k_3/k_t = 0.005$. Brown points correspond to $\beta = 0.95$ and squeezing $k_3/k_t = 0.0025$. The dashed line has approximate form $f_{NL}^{sq} = -0.006 + 0.92f_{NL}^{eq}$, which is common for both squeezings. **Bottom left**: Distribution of the estimated ‘shape running’ parameter $n_{\text{shape}}$. **Top right**: Distribution of percentage shift between $f_{NL}^{eq}$ and $f_{NL}^{sq}$ for $\beta = 0.95$. 

\[ \beta = 0.95 \quad \bullet \quad \beta = 0.9 \]

\[ \text{estimate of shape running } n_{\text{shape}} \]

\[ \% \text{ shift from equilateral to squeezed} \]

\[ -10 -8 -6 -4 -2 0 \]

\[ 0 \quad 200 \quad 400 \quad 600 \]
between 1/3 and 1/2. Meanwhile, $k_3/k_t$ varies between 1/3 and 0.0025. Assuming the strongest dependence comes from the squeezed momentum, the effective ‘shape’ running can be written in terms of a parameter $n_{\text{shape}}$ \cite{206, 208}

$$f_{\text{NL}} \approx f_{\text{NL}}^\star \left( \frac{k_3}{k_t} \right)^{n_{\text{shape}}},$$

(2.52)

where $f_{\text{NL}}^\star$ is a fiducial value taken here to be the value of $f_{\text{NL}}$ on an equilateral configuration at $k_t = 3k_\star$. In the bottom left panel of Fig. 2.15 we plot the distribution of $n_{\text{shape}}$ for our sample of squeezed configurations. It shows very pronounced clustering near $n_{\text{shape}} \approx 0.02$. The scale dependence of the spectrum is divided out of $f_{\text{NL}}$ by construction, but it is possible that this shape dependence is generated by the same underlying process \cite{208, 209}. In particular, the shape of the $n_s$ distribution in Fig. 2.3 is qualitatively similar to (but not the same as) the distribution of $n_{\text{shape}}$ in Fig. 2.15. One might therefore expect the shape running to be small when the spectrum falls in the observationally-allowed window, but our ‘squeezed’ catalogue contains too few trajectories to make a definitive statement. It would be interesting to study the shape running in a larger sample.

### 2.3.2.4 Large non-Gaussianity from rapid change of angular minima

In certain very rare examples we observe the synthesis of large non-Gaussianity, apparently caused by abrupt shifts in the brane trajectory. These occurrence of these rare shifts was recognized by Agarwal et al., who conjectured they might generate appreciable three-point correlations \cite{146}. Our results demonstrate this conjecture to be essentially correct.

**Typical evolution.**—In Fig. 2.16 we show the time evolution for a typical trajectory exhibiting a rapid transition between distinct angular minima. In the top-left panel we plot the time evolution of the background fields. The transition occurs at roughly $N \approx 6.8$ e-folds from the initial time (marked by the dotted vertical line) and is characterized by rapid evolution of $\phi_2$ and $\psi$. The motion is overdamped by Hubble friction and the fields settle smoothly into the new vacuum. After the transition the system approaches an adiabatic limit. The dashed vertical line indicates the time at which maximum $|f_{\text{NL}}|$ is attained.

In the bottom-left panel we show the time evolution of $f_{\text{NL}}^{\text{eq}}$, evaluated on the scale $k_t/3 = k_\star = 0.002 \text{Mpc}^{-1}$. The amplitudes $f_{\text{NL}}^{\text{old}}$ and $f_{\text{NL}}^{\text{eq}}$ exhibit qualitatively similar time-dependence, although there are differences of detail. In each of these measures, the amplitude of 3-point correlations exhibits rapid growth after the transition up to a peak at $N \approx 9.6$, followed by rapid decay. Eventually, as the adiabatic limit is reached, the equilateral amplitude asymptotes to $f_{\text{NL}}^{\text{eq}} \sim 15.8$, with slightly different values for the other configurations. We plot the evolution only up to $N = 20$, beyond which the dynamics become uninteresting. For ease of comparison the top-left and bottom-left panels share the same horizontal scale.
Figure 2.16: Example trajectory exhibiting rapid transition between angular minima. **Top left:** Evolution of background fields. The dotted vertical line indicates the approximate time of the transition. The dashed vertical line indicates the time at which peak $|f_{\text{eq}}^{NL}|$ is achieved (see bottom left panel). **Bottom left:** Evolution of $f_{\text{eq}}^{NL}$ on a scale $k_t/3 = k_* = 0.002\text{ Mpc}^{-1}$. The axes match the top left panel, and the approximate transition time and peak $|f_{\text{eq}}^{NL}|$ are marked in the same way. **Top right:** Time evolution of the dimensionless 3-point functions $(k_1 k_2 k_3)^2 \langle O_1(k_1) O_2(k_3) O_3(k_3) \rangle'$, for selected operators $O_i$. The prime $'$ attached to the correlation function denotes that the momentum-conserving $\delta$-function $(2\pi)^3 \delta(k_1 + k_2 + k_3)$ has been removed.
This time evolution of $f_{\text{NL}}^\text{eq}$ closely matches the behaviour described by Elliston et al., associated with rolling between two approximate critical points of the potential [119]. In this case the critical points should be regarded as the angular minima connected by the rapid transition. Note that Ref. [119] considered only the case of two-field dynamics in which the trajectory rolls from the vicinity of a critical point at which the Hessian has one dominant negative eigenvalue to the vicinity of a different critical point at which the Hessian has one dominant positive eigenvalue. However, it is plausible that similar behaviour occurs in transitions between critical points with different numbers of positive and negative eigenvalues. This example suggests it may be profitable to develop a general theory of such transitions. For example, the above analysis clearly invites a connexion with the language and methods of Morse theory. To our knowledge, however, this connexion is currently unexplored.

**Numerical fidelity.**—The abrupt character of the transition requires that we exercise caution before accepting numerical results. We have verified that the solutions are stable to adjustment of the time-sample mesh, numerical tolerances, and changes of stepper. For these trajectories we have compared the output from our Dormand–Prince 4th/5th-order stepper with an adaptive Bulirsch–Stoer stepper that can work up to 8th order where the solution has rapid changes or discontinuities. So far as we can judge, our numerical methods appear to correctly control the solution for each correlation function during the transition. Further, despite the abrupt evolution of the background fields, we have verified that the eigenvalues of the mass matrix evolve smoothly.

In the top-right panel of Fig. 2.16 we plot the time evolution for a representative sample of field-space correlation functions. These demonstrate that the evolution is smooth and the crucial decay during the subhorizon era is being followed accurately. It is this era that is typically most difficult to integrate, and conversely noise in this region is usually a reliable indicator of numerical problems [79, 75]. For this trajectory there are no significant subhorizon oscillations. Other trajectories do exhibit them in off-diagonal correlators (12) as a consequence of unbalanced phase factors $\sim e^{i(E_1-E_2)t}$ between states 1 and 2 with energies $E_1, E_2$. Where we have performed spot checks on such trajectories, the solutions are stable to the changes described above and the oscillations appear to be smoothly resolved.

**Summary.**—We caution that trajectories exhibiting these rapid transitions are rare; there is no sense in which they can be regarded as typical features of the D3/\overline{D3} model—although, curiously, the single adiabatic trajectory satisfying the (separate) WMAP7 3σ bounds on $A_s$ and $n_s$ is of this type. We observe ten examples of this kind in our primary catalogue for which $|f_{\text{NL}}^\text{eq}|$ is larger than unity, although there are certainly more examples that fall below this arbitrary threshold. The cases where $|f_{\text{NL}}^\text{eq}| > 1$ correspond to $f_{\text{NL}}$ equal to $-18.7, -12.3, -9.3, -8.71, -8.49, -7.31, 11.0, 12.3, 15.8, 75.8$. Their rarity suggests it is unlikely that observable signatures of these transitions could be used to constrain the model. Nevertheless, despite their atypicality, these trajectories are very
interesting as examples of behaviour that has been observed in toy models [119], but not (to our knowledge) in a model motivated by a purpose other than its ability to generate large 3-point correlations.

### 2.3.3 ‘Small’ ensemble comparisons

We now consider the ‘small’ ensembles produced by the PyTransport pipeline. (See Table 2.7.) In general these show that observational predictions are strikingly insensitive to many of the arbitrary choices made in §2.1.2. Specifically, we find very little difference between our principal catalogue and the ‘small’ catalogues produced by varying initial conditions, $\phi_{\text{UV}}$, or the truncation point of the potential. (When varying the initial conditions we do observe a preferred non-zero initial $x$-velocity of the brane, but this does not propagate into the final distributions.)

**Homogeneous model.**—Significant changes do occur when dropping contributions to the potential sourced by the bulk flux product $g_s|\Lambda|^2/96$. We describe this as the ‘homogeneous’ case (see Table 2.7), which corresponds to setting $\mathcal{C}_{LM} = 0$. This is an artificial test in the sense that there is no meaningful limit that would justify dropping the $\mathcal{C}_{LM}$ while retaining the zero modes of the bulk Laplacian. However, the outcome is still interesting because it illuminates how features of the model arise from particular sectors of the potential.

In Fig. 2.18 we plot the mass distribution for this catalogue, measured at 55 e-folds prior to the end of inflation. In comparison with Fig. 2.6 the heavy eigenstates are more nearly degenerate, and their numerical values of $m^2/H^2$ are significantly smaller. Meanwhile, the tachyonic state (red distribution in Figs. 2.6 and 2.18) has a significantly broader tail towards negative values. The conclusion is apparently that the heavy masses are dominated by contributions sourced from the bulk fluxes. It is this feature that causes us to speculate that the difference in mass spectra reported by Agarwal et al. [146], Dias et al. [107] and this work may be associated with differing treatment of these sourced modes.

**Figure 2.17:** Change in observables for the ‘homogeneous’ catalogue. No cut is applied to $A_s$. **Left:** distribution of $f_{\text{NL}}^\text{eq}$ with the full potential (blue) and in the homogeneous case (red). **Right:** distribution of $n_s$ with the same conventions.
Because the sourced modes also change the distribution of the lightest mass eigenstate, it is possible that they contribute to the difference in $\epsilon$ observed between all three studies.

By itself, a change in the spectrum of heavy modes need not imply any shift in the distribution of observables. Therefore in Fig. 2.17 we plot the distributions for $f_{NL}^{eq}$ and $n_s$ in the homogeneous case, without applying a cut on $A_s$. Although the peak of each distribution remains fixed, the shape and length of the tail is adjusted significantly. We conclude that the impact on observables is modest. However, if treatment of the flux-sourced modes is really responsible for the differences observed between our analysis and Agarwal et al. or Dias et al., it is possible that conclusions at the level of observables may exhibit only limited sensitivity to these differences.

$\alpha = 0$ model.—The last study is the restriction to $\alpha = 0$, for which the posterior $Q$-distribution was already discussed in Fig. 2.8. This enables us to connect our analysis with the previous studies reported by Agarwal et al., Dias et al. and McAllister et al., each of which set $\alpha = 0$. We do not observe significant shifts in the distribution of observables, except for a small shift in the $A_s$ distribution (see Fig. 2.19). This conclusion was anticipated by Hertog & Janssen [155]. On this basis, it appears that the differences observed between our analysis and these previous studies should be attributed to structural differences in the potential rather than changes in the sampling procedure.

### 2.4 Discussion

In this paper we have revisited the ‘delicate’ D3/$\overline{\text{D3}}$ brane inflationary model, and computed the amplitude of three-point correlations for the density perturbations. These have not previously been reported. We find that $f_{NL}^{eq}$, $f_{NL}^{old}$ and $f_{NL}^{sq}$ are all highly correlated with nearly degenerate amplitudes close to Maldacena’s single-field prediction $f_{NL} \approx -(5/12)(n_s - 1)$ for $|n_s| \ll 1$ [57], even when $|f_{NL}|$ is large. For trajectories that yield observationally acceptable CMB angular spectra $C_{\ell}^{XY}$, it follows that the ampli-
The amplitude of three-point correlations is usually negligible. For example, the best-fit trajectory for the Planck2015 $TT+TE+EE$ likelihood yields $|f_{\text{NL}}| \sim 5 \times 10^{-3}$. If the restriction to the observationally acceptable region is dropped, the model has a tendency to produce scale-dependent, nonmonotonic power spectra with corresponding $|f_{\text{NL}}|$ of order unity.

A small number of trajectories yield atypically large $|f_{\text{NL}}|$, associated with abrupt transitions of the angular fields between different minima. At the extreme, we observe values $|f_{\text{NL}}| = O(100)$, although still with unacceptable spectral behaviour. Closer to the allowed observational window we find one (adiabatic) trajectory of this type that yields $f_{\text{NL}} = -0.749$ and $n_s = 0.94$ with an acceptable $A_s$. However, because the spectrum is scale dependent the fit is not as good as these values would suggest. The occurrence of large three-point correlation amplitudes for these trajectories was originally conjectured by Agarwal et al. [146]. The large bispectra observed in our catalogue all appear to be produced by such transitions and not ‘gelaton’ or ‘QSFI’ effects [207, 152, 153]. In our smaller catalogue of squeezed configurations we observe some significant shifts between squeezed and equilateral configurations that could possibly be associated with QSFI behaviour, although the corresponding bispectrum amplitudes do not appear to be large. To decide one way or the other would require full computations of the bispectrum shape on these trajectories. Overall, the shape usually appears close to local, although there are exceptional cases that would be interesting to explore.

We find only weak dependence on most of the arbitrary choices made during construction of the potential. In particular, we find our observable distributions are robust to variations in initial conditions and the field-space diameter $\phi_{\text{UV}}$. We find they are also robust to changes in the truncation point of the potential, even for observables derived from the three-point function. This implies that the intrinsic three-body couplings do not play a significant role in setting the amplitude of the bispectrum, which is consistent with the suggestion that the observed large bispectra are sourced by field-space evolution on superhorizon scales associated with transitions between minima. When contributions
to the potential sourced by bulk fluxes are dropped we find that the mass distribution changes significantly, but the effect on observables is much more modest. Variation in the treatment of these contributions may be responsible for the observed differences between independent analyses of the model. If so, there is some basis for optimism that predictions for observables (except perhaps $r$) might be robust.

It is sometimes suggested that $|f_{NL}| \sim 1$ is a generic prediction of inflationary models in which multiple fields are active [130]. This does appear to be the case for what we could call ‘type 1’ models where the fields are spectators during inflation, but become active in the post-inflationary universe—as in the curvaton, inhomogeneous-end-of-inflation and modulated reheating models [210]. Meanwhile, there is relatively little evidence to guide our expectations for ‘type 2’ models in which the fields are active during inflation, but the evolution becomes adiabatic before reheating. Many toy models of this type produce negligible $|f_{NL}|$. The D3/D3 model represents a more realistic example of this type, which can be attributed to the frequent emergence of an effectively adiabatic trajectory long before observable scales leave the horizon. To set our expectations for the interpretation of upcoming large galaxy surveys [211], it would be exceptionally interesting to know whether the same conclusion extends to a larger class of multiple-field models motivated by ideas in high energy physics.

In common with all prior analyses, we find the model is not favourable observationally. Our best-fit trajectory yields a log-likelihood $\ln L \approx -600$ based on Planck2015 temperature and polarization measurements [185], but most trajectories produce spectra that are significantly too scale dependent—even where the amplitude falls in the allowed range. Typically the bispectrum will also be scale dependent whenever the spectrum is scale dependent. Additionally, there is evidence for a characteristic shape running $n_{\text{shape}} \approx 0.02$ in the bispectrum that may possibly be generated from the same source as the scale dependence of the power spectrum. It is not yet clear whether the shape running can be significant when the scale dependence of the bispectrum is small—but, if so, this could be an interesting observational signature.

Our software pipeline depended on development of a new interface between CppTransport and CosmoSIS [129]. A future release of CppTransport is scheduled to include this interface, which can be used with any inflationary model. A similar interface to the more recent Cobaya framework [212] is also planned. These interfaces dramatically simplify the construction of end-to-end pipelines to constrain inflationary models directly from data—a capability which has been available for a long time in the collider phenomenology community, but which has taken longer to become mainstream for inflationary model analysis.

To mitigate the risk of implementation errors we deploy two independent pipelines. These show excellent agreement at the level of individual trajectories, but disagree in detail at the level of the entire catalogue. The factors driving this disagreement have not yet been identified, but we believe they relate to the sampling implementation. After applying the cut $A_s < 10^{-4}$ the catalogues are brought into statistical agreement. Because our
codes operate at tree-level, a cut of this kind would likely be required anyway to remove trajectories on which there is a risk of the leading loop corrections becoming relevant.

Despite the complexity of the model, compute times are not prohibitive. Samples of the two-point function can be computed in a time of order seconds. They could perhaps be included in a parameter estimation Monte Carlo if not too many samples are needed to correctly predict the shape of the spectrum. Samples of the three-point function are substantially more expensive, with integration times of the order of a few hundred seconds depending on the underlying hardware. Therefore inclusion of this information in a Monte Carlo is impractical for the D3/D3 model. In simpler models the compute time is much smaller, although even in this more optimistic case it would be necessary to sample the bispectrum sparsely rather than compute its full shape. It is an interesting question whether reliable methods can be developed to incorporate this information in a practical likelihood calculation.

The D3/D3 model brings other computational challenges. We find that observables derived from the three-point function can require enhanced numerical precision for accurate computation, which is presumably a consequence of roundoff error due to large cancellations between the many intermediate terms that appear in this model. A second example is accurate computation of the tensor spectral index \( n_t \), because the tensor spectrum is extremely flat. Nevertheless, such accurate computations seem worth pursuing because of the possibility that rare trajectories can violate the ‘consistency bound’ \( r \leq -8n_t \) due to significant anticorrelations imprinted at horizon exit.

**Comparison to previous work.**—Our results are in qualitative agreement with previous analyses of the model \([146, 107, 151]\). However there are quantitative differences. In particular there is evidence for some systematic differences between the trajectories in our catalogue and those constructed by Agarwal et al. and McAllister et al. Our typical values of \( Q \) and \( \epsilon \) (and hence \( r \)) are somewhat different, and our typical mass spectra have a qualitatively different structure but similar numerical magnitudes. There is also some evidence, although less strong, for similar systematic differences relative to Dias et al. We have suggested these differences may be related to a different treatment of contributions to the potential sourced by bulk fluxes. There is also our inclusion of \( \alpha \) as a sampling parameter, whereas Agarwal et al., Dias et al., and McAllister et al. imposed \( \alpha = 0 \). This seems to be responsible for some of the shift in the (unobservable) parameter \( Q \), and also slightly broadening the distribution of \( A_s \). Our typical values of \( \epsilon \) and \( r \) are relatively close to those reported to Dias et al., although not the same. However, conversely, our mass spectra are in closer agreement with McAllister et al. We find a population of trajectories yielding red values for \( n_s \) that was not identified by previous analyses. The observational relevance of this population is not yet clear because their spectra almost certainly contain interesting features.

Despite these differences, there are many areas of agreement between our results and
those reported by earlier studies. The overall shape of the mass spectrum, the form (but not normalization) of the $r$-vs-$n_s$ correlation, the occurrence of rapid transits between different minima of the angular fields, the presence of strong anticorrelation between field fluctuations at horizon exit (‘destructive interference’ in the language of McAllister et al.) and possible transgressions of the ‘consistency bound’ are all points of agreement. Although not discussed explicitly in §2.3, our catalogue reproduces a similar probability of inflation and distribution of total number of e-folds $N_{\text{tot}}$ to those reported by Agarwal et al. However, Agarwal et al. gave a persuasive analytic argument for the functional form of the distribution of $N_{\text{tot}}$, which suggests it does not depend sensitively on properties of the potential.

A natural question is whether our conclusions are reliable given the apparent discrepancies between different analyses of the model, which we summarise in Table 2.8. Ultimately this will require forensic comparison between the separate implementations of the D3-brane potential. The most optimistic outlook is that these differences reflect only minor divergences in construction of the potential, against which observables may be fairly robust, as in §2.3.3. The broad qualitative agreement between all three analyses apparently does suggest there can not be very significant differences in the structure of the D3-brane potential, at least when it is able to support an extended epoch of inflation.

To assist future comparison we have attempted to give enough detail in §2.1 that it will be possible for third parties to replicate our analysis. Also, we have made our model specification and sampling parameters available to download from the Zenodo open data repository. These files are provided under a permissive CreativeCommons attribution license.

This situation illustrates the advantages of a standardized format for specification of inflationary models, analogous to the Universal FeynRules Output format used in the collider phenomenology community to specify the particle content, coupling constants and interactions of a model [104]. As inflationary models become more complex, there seems a persuasive argument for the community to converge on a standardized way to exchange similar specifications.
| Quantity          | Marzouk et al. (MMS)                                                                                                                                                                                                 | Dias et al. (DFL)                                                                                                                                                                                                 | McAllister et al. (MRX)                                                                                                                                                                                                 | Agarwal et al. (ABMX)                                                                                                                                                                                                 |
|------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Adiabicity       | Under stringent and relaxed conditions, 64% and 95% of samples were found to be adiabatic. Criteria based on $M_{10}^2$ eigenvalues.                                                                                | All samples were measured as adiabatic. Criteria based on phase-space focussing.                                                                                                                                                                                        | Samples generally reported as adiabatic. Criteria Based on relative power between entropic modes having $10^{-10}$ less power than the adiabatic power.                                                               | –                                                                                                                                                                                                                   |
| Mass spectra     | $H_{\text{exit}}$ masses evenly spaced, with lightest mode extending tachyonically. Heaviest 5 modes remain approximately stable as a function of $N$, with further tachyonic drift of lightest.                                         | $H_{\text{exit}}$ masses qualitatively similar to MMS. Lightest mode matches well, heavy masses $\mathcal{O}(2) \times$ larger.                                                                                                                                  | $H_{\text{exit}}$ masses qualitatively similar to MMS in magnitude, with lightest state similarly distributed. 2nd and 3rd of the heavy states are highly degenerate, as are the 4th and 5th.                                                                  | –                                                                                                                                                                                                                   |
| $\epsilon$      | Bimodal distribution evaluated at $N_{\text{end}} - 60$, with primary (secondary) mode at $10^{-10}$ ($10^{-4}$). 50% (30%) of samples exceeded $\epsilon > 10^{-8}$ ($10^{-6}$).                                              | –                                                                                                                                                                                                                                                                     | Distribution centered at $10^{-12}$ with unspecified evaluation time.                                                                                               | Samples yielded $10^{-15} \lesssim \epsilon \lesssim 10^{-11}$ when evaluated at $N_{\text{end}} - 60$.                                                                                                 |
| $Q$              | Successful realisations closely clustered about $Q \sim 0.006$.                                                                                                                                                      | Explored as hyperparameter, and fixed $Q = 1.4$ for subsequent analysis.                                                                                                                                                                                                 | Potential configuration reported to be same as ABMX.                                                                                                                                                                                                                   | Successful realisation closely clustered about $Q \sim 0.04$.                                                                                                                                                        |
| $n_s$ vs. $r$    | Most trajectories cluster near $n_s = 2$, with most sample points bounded between $1 \lesssim n_s \lesssim 2.5$ and $-4.6 < \log_{10} r < -7.0$.                                                                   | From Fig. 17 of [107], we infer $-9 \lesssim \log_{10} r \lesssim 7$ and $0.9 \lesssim n_s \lesssim 1.4$.                                                                                                                                                           | –                                                                                                                                                                                                                   | Samples roughly bounded between $0.83 \lesssim n_s \lesssim 1.10$ and $-14 \lesssim \log_{10} r \lesssim -11$.                                                                                                 |

**Table 2.8:** Comparative summary between different D3-brane analyses.
Chapter 3

Constraints on $\tau_{NL}$ from Planck temperature and polarization

Kareem Marzouk, Antony Lewis and Julien Carron  JCAP08(2022)015

A wide range of observations appear to be broadly consistent with a Universe that started
the hot Big Bang phase with purely Gaussian adiabatic super-horizon scalar perturbations
in an otherwise Friedmann-Robertson-Walker background. Gaussian adiabatic perturba-
tions have only one locally varying degree of freedom (‘single clock’), which is not locally
observable on super-horizon scales since it corresponds to a local re-definition of the time
coordinate or scale factor [57, 213, 214]. Gaussian adiabatic models are simple and highly
predictive, for example, the small-scale CMB perturbations in each independent Hubble
patch at recombination are expected to be statistically identical. This would be consistent
with inflation, where perturbation generation is dominated by the effect of a single effec-
tive scalar field. Any deviation from this simple prediction would be powerful evidence
for a more complex multi-field inflationary evolution, or other new early universe physics
involving more than one local degree of freedom. In this paper we constrain large-scale
modulation of small-scale CMB fluctuations by reconstructing the modulation field using
the latest Planck data.

In the language of primordial non-Gaussianity, a large-scale modulation of small-scale
power corresponds to ‘squeezed’-shape non-Gaussianity (see Ref. [102, 215, 216] for re-
views). We model the initial non-Gaussian curvature perturbation $\zeta$ as a locally-modulated
version of an underlying Gaussian field $\zeta_0$, so that in real space

$$\zeta(x) \approx \zeta_0(x)(1 + \phi(x)),$$

where $\phi$ is some modulating field, and we assume $|\phi(x)| \ll 1$. If $\phi$ is correlated to $\zeta_0$, there
will be a non-zero local (squeezed) curvature bispectrum, as commonly parameterized by
$f_{NL}$, corresponding in the CMB to the amplitude of the small-scale CMB power spectrum
varying in a way that’s correlated to the large-scale temperature. However, whether or not
$\phi$ is correlated to $\zeta_0$, there will always be a non-zero squeezed trispectrum, corresponding to
the small-scale $\zeta$ power varying spatially with $\phi$. In the inflationary context this is usually
parameterized by $\tau_{NL}$, where $\tau_{NL} \geq \left( \frac{5}{6} f_{NL} \right)^2$ depending on the degree of correlation of $\phi$
and $\zeta_0$ [59]. The local bispectrum constraint $f_{NL} = -0.9 \pm 5.1$ from Planck [27] means that
any $\tau_{NL}$ from a nearly scale-invariant $\phi \propto \zeta_0$ must be very small, however there remains
the possibility of an uncorrelated modulation giving rise to a more substantial $\tau_{\text{NL}}$.

The best current empirical constraint comes from the Planck 2013 nominal-mission temperature data, $\tau_{\text{NL}} < 2800$ (95% CL) [217], where $\tau_{\text{NL}} \sim 500$ would correspond to a modulation at the $|\phi| \sim 10^{-3}$ level.\textsuperscript{1,2} This result was a conservative upper limit, accounting for evidence of frequency-dependent foregrounds or systematic modulation in power (particularly in the octopole). With full-mission Planck data [219] the temperature noise is substantially reduced and the CMB polarization data give additional constraining power; combined with better cleaning of foregrounds and modelling of systematics, it should therefore be possible to substantially improve on the first Planck constraint. The updated NPIPE analysis pipeline [220] is also now available using slightly more data, lowering the noise further, as well as offering the prospect of a better control of systematics and more reliable simulations.

In practice, only large-scale modes of the modulation field can be constrained well, since small-scale variations in power are impossible to distinguish from random fluctuations in local realizations of the modes [221, 222]. We focus on modulation modes corresponding to multipoles $L \lesssim 100$, where at recombination the modulation field remains largely super-horizon and approximately constant through the last-scattering surface. In this case, a primordial curvature modulation translates directly into a large-scale modulation of the small-scale CMB [222, 217]

$$X(\hat{n}) \approx X_g(\hat{n})[1 + \phi(\hat{n}, r_s)] \equiv X_g(\hat{n})[1 + f(\hat{n})],$$

(3.2)

where $X$ is the small-scale CMB temperature perturbation or polarization field, $X_g$ is the corresponding standard unmodulated linear Gaussian field, and $r_s$ is the radial distance to last scattering. The two-dimensional modulation field $f(\hat{n}) \equiv \phi(\hat{n}, r_s)$ parameterizes the modulation directly in terms of the CMB observables, and can be reconstructed from the data using quadratic estimator techniques [223]. For measuring $\tau_{\text{NL}}$, where most of the signal is in modulation modes $L \ll 100$, the super-horizon approximation is accurate at the percent level, allowing a greatly simplified analysis [222] that is nearly optimal. With no additional computational cost we can also constrain $f(\hat{n})$ more generally, without assuming $\tau_{\text{NL}}$ form, allowing for more general scale dependence of the modulation field. The analysis is then equivalent to looking for a power modulation, as possibly suggested by hints for statistical anisotropy [224, 223, 225, 226]. We shall however assume that all the small-scale modes are modulated as implied by the model of Eq. (3.2), rather than allowing for $X_g$ to be modulated in a different way depending on the CMB harmonic scale. This allows us to place good constraints using the full power of high-resolution Planck data, and avoids issues regarding possible a posteriori choices for more general

\textsuperscript{1}We quote constraints assuming no other non-Gaussian signals apart from CMB lensing (and other effects modelled in the simulations) are present. Ref. [218] find a weaker joint constraint with $g_{\text{NL}}$, but also used a more suboptimal estimator.

\textsuperscript{2}This order of modulation follows from the theory spectrum $C_L^f = \tau_{\text{NL}} C_L^\zeta$; (see later Eq. 3.34). Taking the dominant dipole contribution, we have $C_{L=1}^f \approx 4 \times 10^{-6}$, where the curvature power spectrum is computed theoretically e.g. with CAMB. The characteristic amplitude of harmonic coefficients are then $|a_{1M}| \sim 2 \times 10^{-3}$, such that the corresponding real-space modulation is approximately $10^{-3}$.\textsuperscript{3}
parameterizations (for a discussion of scale-dependent models, see e.g., Ref. [227, 228]).

The signal modulation of Eq. (3.2) is of exactly the same structure as having an optical depth to last scattering ($\Delta \tau$) that varies with line of sight, with

$$X(\hat{n}) = e^{-\Delta \tau(\hat{n})}X_g(\hat{n}) \approx X_g(\hat{n})[1 - \Delta \tau(\hat{n})].$$

(3.3)

The modulation reconstruction can therefore also be used to constrain patchy reionization [229, 230, 231, 232, 233, 234]. The main difference is that the standard $\tau_{NL}$ model imposes a specific scale-dependence to the modulation power spectrum, where the signal of interest is concentrated on very large scales. Optical-depth perturbations by contrast have signal out to small scales, since the modulation more closely follows the density spectrum than the curvature/potential spectrum. Interpretation of the patchy reionization constraint is also more complicated as it depends on the details of the reionization model. There is also a significant correlation with CMB lensing [235, 236], and polarization can also be generated by scattering at reionization. In this work we focus on constraining $\tau_{NL}$, but our full modulation reconstruction power spectrum (which is dominated by the temperature modulations) demonstrates a non-detection which could also be translated into a constraint on specific reionization models. For Planck-level sensitivity, the patchy reionization signal is not expected to significantly contaminate $\tau_{NL}$ at the level that could be detected.

### 3.1 Theory and Methodology

#### 3.1.1 A squeezed $\tau_{NL}$ trispectrum

We construct the general trispectrum from four wavevectors $l_1, l_2, l_3, l_4$ that define a closed quadrilateral in harmonic space. We will consider $X = X(I)$ to be a random field, which may correspond to, for example, the temperature fluctuations of the CMB at last scattering. The four-point function $\langle X(l_1)X(l_2)X(l_3)X(l_4) \rangle$ will receive two contributions, which are referred to as “connected” and “disconnected” pieces. Regardless of the statistical properties of the field $X$, the disconnected component can be factorized into the products of two-point function $\langle X(l_i)X(l_j) \rangle$, i.e., power spectra, which are typically non-zero even if $X$ is a Gaussian random field. By comparison, the connected part of the four-point function is exactly zero when $X$ is Gaussian. We can therefore probe the non-Gaussianity of the field $X$ by analysing the connected trispectrum.

In this work, we consider a diagonally squeezed form of the trispectrum. To understand this, let us first consider a triangle in harmonic space. If we take this triangle, and squeeze one of its sides such that $|L| \ll |l_1|, |l_2|$ and $|L| \approx |l_1 + l_2|$, then we necessarily have one large-scale mode and two small-scale modes. If the bispectrum is positive in regions of space with a positive large-scale mode, the small-scale power over those regions must be enhanced; correspondingly, it must be suppressed if the bispectrum is negative; and
Constraints on $\tau_{NL}$ from Planck temperature and polarization

![Diagram of trispectrum with wavevectors](image)

**Figure 3.1:** *Left:* Example trispectrum. Wavevectors form a closed quadrilateral in harmonic space. The quadrilateral can be described equivalently by two triangles related by a common wavevector $L$. *Right:* Schematic of a squeezed trispectrum, corresponding to the $\tau_{NL}$ signal peak; $L \ll l_1, l_2, l_3, l_4$ and $L \approx |l_1 + l_2| \approx |l_3 + l_4|$.

vice versa for a negative large-scale mode. The small-scale power therefore appears to be modulated by the large-scale mode, with a correlation between the small-scale power and the large-scale mode.

By extension to this triangular picture, we could append two further modes $(l_3, l_4)$ to the squeezed side of the triangle to form a closed quadrilateral. So long as these additional modes are much longer than the common diagonal length, $|L| \ll |l_3|, |l_4|$, then we will form a diagonal-squeezed trispectrum. We display this geometry in Figure 3.1. In this case, the field with large-scale mode $L$ is not one of the observed modes defining the measured trispectrum, and does not have to be directly observable itself. If there is a correlation between the large-scale mode and the curvature perturbation, there will potentially be an observable bispectrum. However, the large-scale mode may be uncorrelated, or weakly correlated, in which case there may be a substantial trispectrum without a corresponding observable bispectrum.

If the large-scale mode is associated with a scalar field, as expected for example in simple multi-field inflation models, the squeezed-diagonal trispectrum is independent of the relative orientation of the small and large-scale modes. The small-scale power is therefore expected to remain locally isotropic, but to have a spatial modulation determined by the large-scale mode. Since the large-scale mode may not be observable, the connected trispectrum is what quantifies this observable modulation in small-scale power.

This squeezed diagonal trispectrum is what is known as the $\tau_{NL}$-shape. Its definition is usually accompanied with another shape constructed via squeezing one outer length of the quadrilateral such that $|l_1| \ll |l_2|, |l_3|, |l_4|$, rather than across its diagonal. This $g_{NL}$-form trispectrum produces a large-scale modulation of a small-scale bispectrum [102].

Together, we conventionally write the connected trispectrum as the sum

$$\langle X(l_1)X(l_2)X(l_3)X(l_4) \rangle_c = \langle X(l_1)X(l_2)X(l_3)X(l_4) \rangle_{\tau_{NL}} + \langle X(l_1)X(l_2)X(l_3)X(l_4) \rangle_{g_{NL}}.$$

(3.4)

The $\tau_{NL}$ and $g_{NL}$ trispectra can be produced by local quadratic and cubic corrections to an otherwise Gaussian initial perturbation field. It is corrections of this type that are
computed in the $\delta N$ formalism for inflationary perturbations [58, 227, 215, 100], and these are often used as key predictors for models of inflation described by multiple fields. The fact that the peak signals for $\tau_{NL}$ and $g_{NL}$ are attained in different momentum configurations allow us to form constraints almost independently of one another. In the remainder of this paper, we will only consider the $\tau_{NL}$ trispectrum, and refer the reader to the 2018 constraints on $g_{NL}$ by the Planck collaboration [27]. It is usually also assumed that the large-scale modes are approximately scale invariant (have a power spectrum proportional to $P_{\zeta}^0$, as would be the case for local $\zeta_0^2$ and $\zeta_3^3$ corrections), which then determines the overall wavenumber-dependence of the signals.

The explicit form of the $\tau_{NL}$ trispectrum in Eq. (3.4) connects the primordial fluctuations to those on the last scattering surface. We can write the trispectrum in its reduced form (i.e. with redundant symmetries eliminated) as [237]

$$p_{ll'lj}^{l_1l_2}(L) = \tau_{NL} \int dr_1dr_2r_1^2r_2^2 F_L(r_1, r_2) \Xi(r_1, l_1, l_2) \Xi(r_2, l_3, l_4),$$

(3.5)

where $F_L(r_1, r_2) = 4\pi \int d\ln k P_\zeta(k) j_L(kr_1) j_L(kr_2)$ and $\Xi$ contains the radiative transfer functions and further spherical Bessel functions $j_L$ required for projecting onto the CMB surface. Exact results can be obtained numerically via Boltzmann codes, but there are analytic simplifications. In particular at recombination the kernel $F(r_1 = r_\ast, r_2 = r_\ast) = C_L^{\zeta\ast}$ and in the squeezed limit the (approximate) constancy of the large-scale mode $L$ relative to the short scale fluctuations means that $\int drr^2 \Xi(r, l, l) \approx (C_l + C_{l_3})(C_l + C_{l_4})$ [222]. The squeezed $\tau_{NL}$ trispectrum at recombination is therefore approximated as

$$p_{ll'lj}^{l_1l_2}(L) \approx \tau_{NL} C_L^{\zeta\ast}(C_l + C_{l_3})(C_l + C_{l_4}).$$

(3.6)

This is exactly the form of squeezed trispectrum expected from the modulated Gaussian field model of Eq. (3.2), where the modulation field is proportional to the curvature perturbation at recombination, $f \propto \zeta_\ast$.

### 3.1.2 The modulation field estimator

We denote the observed CMB temperature and polarization fields with the block vector

$$\hat{\mathbf{X}} = (\hat{T}, +2 \hat{Q}, -2 \hat{U})^T = (\hat{T}, +2 \hat{P}, -2 \hat{P})^T,$$

(3.7)

where we have adopted the use of “spin indices” to indicate the spin-0 temperature field $\hat{T}$ and spin-2 polarization fields $\pm 2 \hat{P} = \hat{Q} \pm i \hat{U}$ in order to write the fields in a unified notation. CMB fields that appear in bold typeface refer to the vector of pixel-data, and if non-bold refer to their scalar value when evaluated at a specific location on the sky, e.g. $\hat{T}(\hat{n}) \in \mathbf{T}$. $\hat{Q}$ and $\hat{U}$ are the Stokes polarization maps which are obtained at the detector level. Each data component in Eq. (3.7) is modelled as the linear combination of the cosmological signal and noise $\hat{\mathbf{X}} = \mathbf{X} + \mathbf{N}$, each of which have the same block vector structure: $\mathbf{X} = (0 \mathbf{X}, +2 \mathbf{X}, -2 \mathbf{X})^T$, $\mathbf{N} = (0 \mathbf{N}, +2 \mathbf{N}, -2 \mathbf{N})^T$. A large-scale modulation that
is approximately constant through the last-scattering light cone will alter each component of the signal vector in the same way, which we define as a linear real-space operation 

\[ s\mathbf{X}(\mathbf{n}) = [1 + f(\mathbf{n})] s\mathbf{X}_0(\mathbf{n}) \]  

(where the modulation acting on the otherwise unmodulated signal \( \mathbf{X}_0 = (0, +2X_0, -2X_0)^T \)). We then model the observed data as

\[ \hat{\mathbf{X}} = \mathbf{B}\mathbf{X} + \mathbf{N}, \]  

(3.8)

where \( \mathbf{B} \) denotes the operation of the instrumental beam (a convolution in real space) and pixel window function.

Assuming that the unmodulated signal and noise components are reliably approximated as Gaussian random fields, we can use maximum likelihood estimates (MLE) to build estimates of \( f \). We follow the methods developed by Hirata and Seljak [238] (for CMB lensing), and later by Hanson and Lewis [223] (for generalized weak statistical anisotropy), defining the (negative) log-likelihood function

\[ \mathcal{L}[f] = -\ln P(\hat{\mathbf{X}}|f) = \frac{1}{2} \hat{\mathbf{X}}^\dagger \left( \mathbf{C}^{\hat{\mathbf{X}}\hat{\mathbf{X}}} \right)^{-1} \hat{\mathbf{X}} + \frac{1}{2} \ln \det \mathbf{C}^{\hat{\mathbf{X}}\hat{\mathbf{X}}} + \text{const.} \]  

(3.9)

The conditional probability \( P(\hat{\mathbf{X}}|f) \) appearing in the log-likelihood is the standard multivariate Gaussian density with data covariance \( \mathbf{C}^{\hat{\mathbf{X}}\hat{\mathbf{X}}} \), which is not isotropic as it depends on the realization of the statistical anisotropy field \( f \). Note that we have represented the modulation \( f \) as the vector \( f \) (here containing only one item) for notational compatibility with the formalism.

We can find \( f \) corresponding to the maximum-likelihood point by identifying the minimum of Eq. (3.9). This can be computed directly by finding where the functional derivative of the likelihood function is zero, \( \delta \mathcal{L}/\delta f^\dagger = 0 \). As explicitly given in [223], the derivative may be conveniently expressed in terms of the combination

\[ \mathcal{H} = \frac{1}{2} \left[ \mathbf{B}^\dagger \left( \mathbf{C}^{\hat{\mathbf{X}}\hat{\mathbf{X}}} \right)^{-1} \hat{\mathbf{X}}^\dagger \frac{\delta \mathbf{C}^{\hat{\mathbf{X}}\hat{\mathbf{X}}}}{\delta f^\dagger} \mathbf{B}^\dagger \left( \mathbf{C}^{\hat{\mathbf{X}}\hat{\mathbf{X}}} \right)^{-1} \hat{\mathbf{X}} \right], \]  

(3.10)

such that

\[ \frac{\delta \mathcal{L}[f]}{\delta f^\dagger} = \langle \mathcal{H} \rangle - \mathcal{H} = 0. \]  

(3.11)

The expectation value \( \langle \mathcal{H} \rangle \) corresponds to a “mean field”, which comes from the derivative of the determinant term. It can conveniently be thought of as a data simulation average.\(^3\)

Solutions for \( f \) that satisfy Eq. (3.11) can be found iteratively, e.g. with a Newton-Raphson approach. This is particularly efficient for weak sources of anisotropy since \( f \approx 0 \), meaning that a single iteration from \( f = 0 \) will yield a good approximate solution to the anisotropy.

\(^3\)From the identity \( \text{Tr} \mathbf{A} = \langle x^\dagger \mathbf{A} c^{-\dagger} x \rangle \), which holds for a random vector \( x \) with \( \langle xx^\dagger \rangle = \mathbf{C} \) and any (square) matrix \( \mathbf{A} \), we have

\[ 2\langle \mathcal{H} \rangle = \left\langle \hat{\mathbf{X}}^\dagger \left( \mathbf{C}^{\hat{\mathbf{X}}\hat{\mathbf{X}}} \right)^{-1} \frac{\delta \mathbf{C}^{\hat{\mathbf{X}}\hat{\mathbf{X}}}}{\delta f^\dagger} \left( \mathbf{C}^{\hat{\mathbf{X}}\hat{\mathbf{X}}} \right)^{-1} \hat{\mathbf{X}} \right\rangle = \text{Tr} \left[ \left( \mathbf{C}^{\hat{\mathbf{X}}\hat{\mathbf{X}}} \right)^{-1} \frac{\delta \mathbf{C}^{\hat{\mathbf{X}}\hat{\mathbf{X}}}}{\delta f^\dagger} \right] , \]
Explicitly, the first iteration produces the estimator

$$\hat{f} = -\left[\frac{\delta}{\delta f}((\mathcal{H} - \mathcal{H})|_0)^\dagger\right]^{-1}((\mathcal{H}_0 - \mathcal{H}),$$

(3.12)

where the lower index “0” indicates evaluation with $f = 0$, corresponding to the initial step. The functional derivative in Eq. (3.12) can be approximated with the Fisher information

$$\mathcal{F}_0 = \left\langle \frac{\delta}{\delta f}((\mathcal{H}_0) - \mathcal{H}_0)\right\rangle,$$

(3.13)

which provides a natural form of normalization for the likelihood based estimator. The normalized quadratic estimator that we use then takes the form

$$\hat{f} = \mathcal{F}_0^{-1}(\mathcal{H}_0 - \langle \mathcal{H}_0 \rangle),$$

(3.14)

where

$$\mathcal{H}_0 = \frac{1}{2} \sum_{s_1, s_2}^1 \bar{\mathcal{X}}^\dagger \left[\frac{\delta C^{XX}}{\delta f}\right]_{0}^{s_1 s_2} \bar{X},$$

(3.15)

and we have introduced the inverse (co-)variance filtered fields, defined as $\bar{X} \equiv B(\mathcal{C}^{XX})^{-1}|_0 \hat{X}$. An analytic expression for the Fisher normalization is in general non-trivial, so in practice we consider the isotropic full-sky limit of the reconstruction to obtain a diagonal simplification that we then correct for with a Monte Carlo correction. We implement a normalization that is consistent with that given in Refs. [223, 222].

The covariance $C^{XX}$ is (in the general case) a $3 \times 3$ block matrix, whose elements $C_{s_1 s_2} = \langle s_1 X s_2 X^\dagger \rangle$ represent the covariance between the spin fields with fixed anisotropy. We can rewrite the product of matrices as the sum over spin-indices to obtain

$$\mathcal{H}_0 = \frac{1}{2} \sum_{s_1, s_2} \bar{X}^\dagger \left[\frac{\delta C^{s_1 s_2}}{\delta f}\right]_{0}^{s_1 s_2} \bar{X}.$$  

(3.16)

We calculate (up to linear order in modulation) the derivative of the covariance to be

$$\left[\frac{\delta C^{s_1 s_2}(\hat{n}_1, \hat{n}_2)}{\delta f(\hat{n})}\right] = \langle s_1 X(\hat{n}_1) s_2 X^\ast(\hat{n}_2) \rangle (\delta(\hat{n} - \hat{n}_1) + \delta(\hat{n} - \hat{n}_2)),$$

(3.17)

where $\delta(\hat{n} - \hat{n}_i)$ appearing on the right-hand side result from locality of the modulation. By expanding $s_1 X(\hat{n})$ into spherical harmonics Eq. (3.17) can be written in terms of the fiducial theoretical power spectra associated with the spin fields

$$\langle s_1 X(\hat{n}_1) s_2 X^\ast(\hat{n}_2) \rangle = \sum_{\ell_1, m_1}^1 \sum_{\ell_2, m_2}^1 B_{\ell_1} B_{\ell_2} s_1 Y_{\ell_1 m_1}(\hat{n}_1) s_2 Y^\ast_{\ell_2 m_2}(\hat{n}_2) \langle s_1 X_{\ell_1 m_1}(\hat{n}_1) s_2 X^\ast_{\ell_2 m_2}(\hat{n}_2) \rangle,$$

(3.18)

where $B_{\ell}$ is the harmonic space representation of the instrumental beam and pixel window.
The quadratic building block for the modulation estimator is then

\[ S\mathcal{H}_{0}^{XX}(\hat{n}) = \sum_{s \in S} \bar{X}(\hat{n})^{*} s X_{WF}(\hat{n}), \tag{3.19} \]

where \( S \) is the set of CMB field spin indices to sum over. In practice, we implement a version of conjugate-gradient descent to obtain an approximate iterative solution for the Wiener-filtered (WF) fields \( X_{WF} = C\bar{X} \equiv CB^{\dagger}(CBX)^{-1}X \), where \( C \) is the covariance of fiducial power spectra, for both data and the simulations [239, 240]. The inverse-variance filtered (IVF) fields are simply related to the WF fields by dividing the matrix of fiducial spectra \( C \). The quadratic estimator remains unbiased, though slightly suboptimal [241], if an approximate version of \( CXB \) is used, as long as the normalization is computed consistently. For simplicity, we take \( C^{TE} = 0 \) in the theory part of the covariance \( C \), to allow separate filtering of temperature and polarization maps, and approximate the noise as isotropic over the unmasked area. The inverse noise is taken to be zero in the masked regions, so that masked pixels do not contribute to the WF maps. The modulation estimator does respond to \( B \) modes, but for \( Planck \) they contribute negligibly to the \( \tau_{NL} \) constraint.

The equivalent harmonic space representation of (3.19) can be directly obtained via the spin-zero harmonic transformation

\[ S\mathcal{H}_{0,LM} = \int d\hat{n} \ S\mathcal{H}_{0}^{XX}(\hat{n}) Y_{LM}^{*}(\hat{n}). \tag{3.20} \]

By defining the set of spin-indices, \( S \) appearing in Eq. (3.19) we can construct different estimators. For example, we obtain the (unnormalized) temperature estimator given in [223] by setting \( S_{TT} = \{0\} \)

\[ S_{TT} \mathcal{H}_{0}^{XX}(\hat{n}) = \bar{T}(\hat{n}) T_{WF}(\hat{n}), \tag{3.21} \]

whereas the polarization and minimum variance estimators are defined by \( S_{PP} = \{+2, -2\} \) and \( S_{MV} = \{0, +2, -2\} \) respectively. These latter estimators have simple expressions analogous to Eq. (3.21). Each estimator can equivalently be expressed in terms of the CMB harmonics \( \{T_{lm}, E_{lm}, B_{lm}\} \), though appear more cumbersome.

The quadratic estimators can reconstruct the modulation field up to the limits of instrumental noise and the cosmic variance of the CMB fields themselves. These Gaussian sources produce a reconstruction noise, \( \mathcal{N}_{L}^{(0)} \), which corresponds directly to the disconnected trispectrum contributions discussed in 3.1.1, and should be subtracted from any modulation field power spectrum estimate. This can be seen by computing the covariance between two quadratic estimators, which will result in Gaussian terms, composed of products of noise and (theoretical) CMB power spectra. Under idealized conditions, this reconstruction noise is statistically isotropic and equivalent to the isotropic Fisher nor-

\[ ^{5} \text{A scalar primordial non-Gaussianity signal should not produce } B \text{ modes, but a modulation does. In the ultra-squeezed limit in which the modulation approximation is exact, the } B \text{-mode contribution also vanishes. Including the } B \text{ modes is correct for modulation reconstruction, but for futuristic measurements where the } B \text{-mode contribution may become non-negligible due to low delensed cosmic variance, the } B \text{ mode contribution should strictly be dropped when using the modulation model as an approximation to constrain } \tau_{NL}. \]
nalization, such that it is straightforward to account for. We normalize using the full-sky analytic result, but correct our final spectra by a Monte Carlo normalization to account for inaccuracies in the analytic result. The reconstruction noise bias and mean field are also calculated from simulations, as described in the following section. Using harmonic representation of the full-sky analytic Fisher normalization gives us the normalized quadratic estimator

$$\hat{s}_{\mathcal{f}_{LM}} = (s_{F_{0,0}}^{-1})_{LM} s_{H_{0,0}}$$

(3.22)

where we have appended indices to the normalization factor in order to keep track of the estimator properties, which will alter the required normalization in general.

We can generalize the quadratic estimator in the Eq. (3.19) to a “joint estimator”, which combines information from pairs of maps which may correspond to e.g. distinct observations of the CMB. For map pairs $X(\hat{n}), Z(\hat{n})$ we define the unnormalized estimator analogous to (3.19)

$$s_{H_0^{XZ}}(\hat{n}) = \frac{1}{2} \sum_{s \in S} \left[ s \tilde{Z}^q(\hat{n})_s X^{WF}(\hat{n}) + s \tilde{X}^q(\hat{n})_s Z^{WF}(\hat{n}) \right],$$

(3.23)

where the symmetrized mean accounts for the fact that in general the quadratic building block is not symmetric under exchange $X \leftrightarrow Z$. Similarly, we write the symmetrized normalization as

$$s_{A_{LM}^{XZ}} = \frac{1}{2} (s_{F_{0,0}^{XZ}} + s_{F_{0,0}^{ZX}}),$$

(3.24)

to obtain

$$s_{\mathcal{f}_{LM}^{XZ}} = (s_{A_{LM}^{XZ}}^{-1}) s_{H_{0,0}^{XZ}}.$$  

(3.25)

Note that if $X = Z$ these definitions automatically recover the former result of Eq. (3.22).

In the remainder of this paper we will use power spectrum estimates of the modulation field computed from the normalized quadratic estimator(s) described in this section.

### 3.1.3 Power spectrum estimator

Power spectrum estimates of the modulation field are computed from the cross-spectrum of pairs of quadratic estimators $s_i \hat{s}_{LM}^{X_1Z_1}$ and $s_i \hat{s}_{LM}^{X_2Z_2}$. Each of these components require mean field subtraction as given in Eq. (3.11), which in practice is calculated as the average quadratic estimate over simulations with $f = 0$. The mean field serves as a map-level correction to the data that accounts for the main sources of statistical anisotropy that are not primordial. These include Doppler aberration and modulation, anisotropic noise and beam anisotropy, as well as estimator response to masked regions of the sky. The accuracy to which the mean field can be estimated is limited by the number and fidelity of the simulations. The Doppler-induced effects are important for $\tau_{NL}$ since most of their signal is at very low $L$, but have a fixed known direction such that they are expected to be accounted for accurately by the simulations.

Monte Carlo noise on the mean field limits accuracy of the reconstruction, but can on
average be nulled at the power spectrum level by ensuring the mean field subtracted from the data reconstruction on each leg has independent Monte Carlo noise. Since realizations of CMB and instrumental noise are uncorrelated from simulation-to-simulation, this can be achieved straightforwardly by constructing two separate mean field estimates from independent sets of simulations. The mean field (on each component $i \in \{1, 2\}$) is

$$s_i J_{LM}^{X_i Z_i, MF} = \langle s_i J_{LM}^{X_i Z_i} \rangle_{\text{sims}},$$

(3.26)

such that the leading contribution to the modulation power spectrum is given by

$$C_L^{S_1 X_1 Z_1, S_2 X_2 Z_2} = \frac{1}{(2L + 1) f_{\text{sky}}} \sum_M \left( s_1 J_{LM}^{X_1 Z_1} - s_1 J_{LM}^{X_1 Z_1, MF} \right) \left( s_2 J_{LM}^{X_2 Z_2} - s_2 J_{LM}^{X_2 Z_2, MF} \right)^*,$$

(3.27)

If we assume that the simulations reliably capture the expected lensing signal\(^7\) the estimator can then be made unbiased by simply subtracting the mean over simulations with zero modulation $C_L^{S_1 X_1 Z_1, S_2 X_2 Z_2} - \langle C_L^{S_1 X_1 Z_1, S_2 X_2 Z_2} \rangle_{f=0}$. However, we can instead use the optimal perturbative trispectrum estimator that does better than this, by correcting for the realization of the Gaussian power in the data map compared to the mean in the simulations [98, 242, 243, 244]. This realization-dependent correction, $N_{L}^{\text{RD},(0)}$, reduces the variance of the estimator, and also makes it perturbatively self-correcting to misestimates of the Gaussian power in the simulations. In the idealized full sky case this can be done by adding the linear correction

$$C_L^{S_1 X_1 Z_1, S_2 X_2 Z_2} \approx \langle C_L^{S_1 X_1 Z_1, S_2 X_2 Z_2} \rangle_{f=0} - \sum_{ab, l} \frac{\partial}{\partial \tilde{C}_{\ell, \text{expt}}} \left( \tilde{C}_{\ell, \text{expt}} - \tilde{C}_{\ell, \text{expt, fid}} \right),$$

(3.28)

where $C_{\ell, \text{expt}}$ is the CMB power spectrum measured on the data and $C_{\ell, \text{expt, fid}}$ the fiducial model of the data power spectrum (i.e. lensed power spectrum + isotropic noise). The dominant part of the derivative correction comes from the response of $N_{L}^{(0)}$ to the power, so we can approximate $\langle C_L^{S_1 X_1 Z_1, S_2 X_2 Z_2} \rangle_{f=0} \approx N_{L}^{(0)}$ in that term; the first term depends on $N_{L}^{(0)}$ and the lensing 4-point contribution to the modulation power.

Accounting for these bias corrections, our estimator for the modulation power spectrum is given by

$$C_L^{S_1 X_1 Z_1, S_2 X_2 Z_2} = \frac{1}{(2L + 1) f_{\text{sky}}} \sum_M \left( s_1 J_{LM}^{X_1 Z_1} - s_1 J_{LM}^{X_1 Z_1, MF} \right) \left( s_2 J_{LM}^{X_2 Z_2} - s_2 J_{LM}^{X_2 Z_2, MF} \right)^* - \Delta C_L^{(0)} - N_{L}^{\text{RD},(0)},$$

(3.29)

where $\Delta C_L^{(0)} = \langle \tilde{C}_L^{S_1 X_1 Z_1, S_2 X_2 Z_2} \rangle_{f=0} - N_{L}^{(0)}$ is the connected lensing contribution. In the general non-full-sky case of power spectra based on four input maps $\{X_1, Z_1, X_2, Z_2\}$ the

\(^6\)Our use of “$\supset$” in this section implies that whatever appears on the right-hand side is a partial contribution to the full quantity given on the left; which is added to linearly.

\(^7\)The simulations have independent random lensing realizations, so lensing produces no mean field on average, but does have a trispectrum. In this paper we do not attempt a joint reconstruction of the lensing and the modulation.
explicit expressions for these terms are:

\[ \Delta C_{L}^{(0)} = \left( \bar{C}_{L} - \left( C_{L} + \bar{C}_{L} \right) \right)_{i \neq j, f=0} \]  

(3.30)

and

\[ \bar{N}_{L}^{\text{RD}(0)} = \left( \bar{C}_{L} - \left( C_{L} + \bar{C}_{L} \right) \right)_{i \neq j, f=0} \]  

(3.31)

where the indices \( i \) and \( j \) imply the maps are Monte Carlo simulations, and \( d \) the data. In practice we implement the convention \( j = i + 1 \) to obtain uncorrelated Monte Carlo pairs, though in principle further permutations of indices could lower the Monte Carlo noise even further. The power spectra have the same definition as in Eq. (3.27) other than the mean field term, which is set to zero. The exception to this is the first term appearing in Eq. (3.30) when each map carries the same index. This term includes lensing 4-point information and will play a further role in our likelihood analysis, it is computed from individual Monte Carlo simulations as

\[ \bar{N}_{L}^{\text{MC}(0)} = \left( \frac{1}{2} \left( C_{L} \bar{C}_{L} - \left( C_{L} + \bar{C}_{L} \right) \right) \right)_{f=0}. \]  

(3.32)

To visualise the relative amplitude of the noise bias corrections as well as the meanfield and a representative \( \tau_{\text{NL}} \) signal, we plot their respective power spectra in Fig. 3.2.

Finally we introduce a calibration factor \( k_{L} \) to form the final estimate

\[ \hat{C}_{L} = k_{L} \bar{C}_{L} \]  

(3.33)

The calibration factor is derived empirically from with-signal simulations such that the ensemble average reconstruction spectra match the theoretical input \( C_{L}^{\text{Theory}} = k_{L} \bar{C}_{L} \). Note that when the calibration factor multiplies the reconstruction noise spectra we apply the same convention; e.g. \( \bar{N}_{L}^{\text{MC}(0)} = k_{L} \bar{N}_{L}^{\text{MC}(0)} \).

The calibration corresponds to an \( O(1) \) multiplicative correction after appropriate binning and smoothing is applied, though is slightly scale-dependent. The dipole receives the largest correction corresponding to the greatest power loss after masking the sky, and towards smaller scales the power is typically reduced by up to 15%. We assess the calibration procedure in more detail in Section 3.3, and have confirmed that it leads to more accurate \( \tau_{\text{NL}} \) estimates, and does not lead to a false detection when applied to without-signal reconstructions. In the remainder of this paper, spectra that are uncalibrated or calibrated will be indicated with a bar or hat, i.e. \( \hat{C}_{L} \) or \( \bar{C}_{L} \) respectively.
3.1.4 $\tau_{NL}$ estimation and likelihood analysis

The simplest estimate that we can construct for $\tau_{NL}$ corresponds to the modulation power spectrum at a fixed scale $L$, normalized by the curvature power spectrum at recombination

$$\hat{\tau}_{NL}(L) = \frac{\hat{C}_L^f}{\hat{C}_L^{\kappa^*}},$$

(3.34)

where we assume that we know $C_L^{\kappa^*}$ theoretically; and can be readily computed with e.g. CAMB. In general $\hat{C}_L^f = \hat{C}_L^{s_1X_1Z_1, s_2X_2Z_2}$ (i.e. is the full modulation power spectrum estimate) but we have omitted indices for convenience. The calibrated estimates are related to the uncalibrated estimates multiplicatively $\hat{C}_L^f = k_L \hat{C}_L^f$, such that $\hat{\tau}_{NL}(L) = k_L \tau_{NL}(L)$.

We adopt the same bar/hat convention for all quantities in our likelihood analysis to make this distinction clear; including the covariance and Gaussian noise bias. Estimates at each $L$ can be combined using an inverse variance weighting. For small signals, an approximately scale-invariant primordial spectrum, and approximating the reconstruction noise as white over the modulation multipole range $L_{\text{min}} \leq L \leq L_{\text{max}}$ that is used, this gives the approximately-optimal combined estimator \(^{[222]}\)

$$\hat{\tau}_{NL} = \frac{1}{R} \sum_{L=L_{\text{min}}}^{L_{\text{max}}} \frac{2L+1}{L^2(L+1)^2} \frac{\hat{C}_L^f}{\hat{C}_L^{\kappa^*}},$$

(3.35)

where the normalization $R$ accounts for the finite multipole range, $R = L_{\text{min}}^{-2} - (L_{\text{max}} + 1)^{-2}$. Regardless of whether or not there is in fact a primordial modulation field, the reconstruction is expected to produce an approximately Gaussian random field on large scales. In the null case, the reconstruction should correspond to a realization of the disconnected Gaussian noise bias, $\delta_{L}^{(0)}$, with an approximately white power spectrum. If there is a primordial modulation, this too is likely to correspond to an approximately Gaussian random field (e.g., if it is generated by quantum fluctuations early in inflation). If it is of $\tau_{NL}$ form, the spectrum would peak on large scales if the signal-to-noise of

---

**Figure 3.2:** Power spectrum of the modulation mean field, reconstruction noise biases, and lensing bias, for the NPIPE (A, B) analysis. The mean field power is computed as the cross-spectrum of the independent mean field components given in Eq. (3.26). For comparison, the dotted line shows the modulation power expected if $\tau_{NL} = 1500$; i.e. $C_L^f = 1500C_L^{\kappa^*}$.
the modulation is sufficiently large to be distinguishable from Gaussian background. We therefore expect that the estimators formed as a sum over \( \hat{\tau}_{NL}(L) \) are approximately \( \chi^2 \)-distributed. Since the signal peaks at low multipoles, this is a very skewed distribution, with 95% of the signal in \( \hat{\tau}_{NL} \) at \( L \leq 4 \) [222].

To constrain the true value of \( \tau_{NL} \) given the high level of realization dependence, we use an approximate model for the posterior distribution of the set of \( \hat{\tau}_{NL} \) estimates. As pointed out in Ref. [245], the statistical properties of a \( \tau_{NL} \) signal are analogous to those assumed when estimating the statistics of the CMB temperature field. We largely borrow from this “single-field” procedure and construct an approximate log-likelihood function on the cut-sky as described in Ref. [246]

\[
-2 \ln P\left( \left\{ \tau_{NL}(L) \right\} || \hat{\tau}_{NL}(L) \right) \approx \sum_{LL'} \left[ \frac{g(\hat{x}(L)) \hat{N}_{NL}(0)(L)}{\hat{M}^{-1}} \right]_{LL'} \left[ \hat{N}_{NL}(0)(L')g(\hat{x}(L')) \right] + \text{const},
\]

(3.36)

using a restricted multipole range \( L_{\text{min}} \leq L \leq L_{\text{max}} \). In the above \( \hat{\tau}_{NL}(L) \) is defined in Eq. (3.34), and \( \hat{N}_{NL}(0)(L) \) is of the same form but using the reconstruction noise power spectrum in place of the modulation power spectrum (and therefore defines the Gaussian noise bias on \( \tau_{NL} \)). \( \hat{M} \) is the covariance matrix of \( \hat{\tau}_{NL}(L) \) estimates inferred from without-signal simulation reconstructions, and is truncated to the prescribed multipole interval before inversion. To better account for the error incurred by the inverse-covariance due to the finite number of simulations, we rescale the inverse by a constant (Hartlap factor [247]) which leads to a slight broadening on the final posterior distribution. Similarly, we account for potential Monte Carlo errors (i.e. error on the mean field and reconstruction noise bias) by including the additive corrections to the diagonal of the covariance [240]

\[
\sigma_{MC,L}^2 = \left( \frac{2}{N_{\text{MF}}} + \frac{9}{N_{\text{bias}}} \right) \frac{2(N_{NL}(0))^2}{(2L+1)J_{\text{sky}}},
\]

(3.37)

where \( N_{\text{MF}} \) and \( N_{\text{bias}} \) are the number of simulations used to compute the mean field and realization dependent noise bias respectively. The likelihood further relies on the functions

\[
g(x) = \text{sign}(x-1)\sqrt{2(x-\ln(x)-1)}, \quad \text{where} \quad \hat{x}(L) = \frac{\hat{\tau}_{NL}(L) + \hat{N}_{NL}(0)(L)}{\tau_{NL}(L) + \hat{N}_{NL}(0)(L)},
\]

(3.38)

where \( \hat{N}_{NL}(0)(L) = k_L \hat{N}_L^{RD,(0)} / C_L \) accounts for realization dependent Gaussian noise fluctuations in the data.\(^8\) Although we reconstruct the modulation field with general \( L \) dependence, for constraining \( \tau_{NL} \) we follow the standard definition by assuming no scale dependence in \( \tau_{NL} \) itself; \( \tau_{NL}(L) = \tau_{NL} \).

Once we have obtained our posterior distribution of \( \tau_{NL} \) estimates, it is straightforward to construct constraints on \( \tau_{NL} \) by numerically integrating the distribution from \( \tau_{NL} = 0 \) up

\(^8\)An exception to this may be “TTxPP” estimators. In this case, the resulting \( N_{L}^{RD,(0)} \) spectra have a small amplitude, which may be negative for some multipoles. For small \( \tau_{NL} \) the logarithm in Eq. (3.38) will then become undefined. Alternative approximations to the reconstruction noise should then instead be used, e.g. \( N_{NL}(L) \approx \sqrt{(2L+1)J_{\text{sky}}M_{LL}/2} \) which is non-negative by definition.
Constraints on $\tau_{NL}$ from Planck temperature and polarization

3.2 Data and simulations

The nominal Planck 2013 results placed constraints on $\tau_{NL}$ using the 143 and 217 GHz intensity maps to reconstruct the modulation field. The foreground signal was partly mitigated by projecting a dust template from higher frequencies, but the Planck 2013 analysis still showed evidence of important residuals. There is also a large spatial variation of the noise, which can mimic the effect of spatially varying power from $\tau_{NL}$ if it cannot be simulated and corrected for accurately. This noise modulation can be mitigated by correlating different maps with independent noise; in Planck 2013 a large quadrupole was substantially reduced by performing the joint (143, 217) GHz reconstruction. The joint reconstruction is obtained by direct application of the estimators discussed in section 3.1.2, where one copy of the inverse filtered harmonics (e.g., as given in Eq. (3.21)) is replaced with those corresponding to a different map. More generally, the noise from observations at different times and different detectors is also approximately uncorrelated, and we can also use joint reconstruction between different data subsets to reduce sensitivity to the spatially-varying noise power.

Advancing from the 2013 to 2018, the data products are derived from a longer observation period via repeated scans of the sky, leading to a lower effective noise in all frequency channels. In addition to this, the release of polarization data enables us to expand the range of estimators at our disposal; from temperature only (TT), to polarization (PP) and minimum variance (MV) combinations, as defined in Section 3.1.2. PP estimators typically offer weak constraining power in comparison to TT and MV, since the small-scale signal is dominated by noise for Planck. However, the MV estimate (which combines information from polarization with temperature), by definition reduces the reconstruction noise when compared to TT and PP (if implemented optimally), and we use it for our best constraints. Other cross-power spectrum estimators can be built by considering the cross between estimators of different types, such as “TT×PP”. Whilst this estimator has low reconstruction noise due to the small noise cross-power spectrum between the temperature and polarization fields, this comes at the expense of a larger estimator variance such that the resulting spectra provide relatively weak constraints.

To see the impact of the changes since the 2013 result, we first produce constraints from updated 143 GHz and 217 GHz temperature-only data, to illustrate improvements associated with lowering the effective noise, whilst remaining subject to the same frequency-dependent signals as in 2013. Secondly, we then use the component-separated SMICA maps in temperature and polarization to reduce foreground contamination, using the joint reconstruction between observations in first and second half of the observational period (HM1, HM2) to mitigate noise modelling uncertainty. Comparison between temperature recon-
Constraints on $\tau_{\text{NL}}$ from Planck temperature and polarization

3.2

The Full Focal Plane (FFP) simulations \cite{248, 249} aim to replicate the statistics of the observed CMB sky. The end products from the map-making process are a set of lensed scalar realizations of the cosmological CMB along with simulated noise + residual systematics.

The CMB simulations rely on fiducial power spectra from which scalar unlensed CMB realizations are constructed. The unlensed CMBs are then lensed by Gaussian realizations of a lensing potential that are independent between simulations. The input cosmological parameters used in the process of generating and subsequently lensing the scalar maps are based on Planck public release (PR) data. For example, in the case of $\text{FFP10}$, which are the set of CMB and noise simulations that accompany the Planck PR3 data, the CMB simulations assume a base $\Lambda$CDM cosmology with parameters inferred from previous PR2 \cite{185} data. Note that the PR3 data are also referred to as the "DX12" data - the latter of which usually appears in technical documentation - and we use these interchangeably throughout this work. The simulated CMB signal contains additional frequency-dependent sources. These include Rayleigh scattering, a second order dipole-induced quadrupole \cite{250}, and a Doppler modulation and (frequency-independent) aberration \cite{251}. Once the lensed scalar maps are generated, they are convolved with $\text{FEBCoP}$ \cite{252}, which accounts for realistic beaming and instrumental pointing.

In order to generate the accompanying $\text{FFP10}$ noise maps a fiducial realization of the scalar CMB is first constructed. The process further relies on sophisticated foreground modelling, via the Planck Sky Model \cite{253}, which contains galactic thermal dust, spinning dust, synchrotron, free-free scattering and CO line emission based on data-based models, as well as random realizations of CIB and SZ that are (unrealistically) uncorrelated to the fiducial CMB lensing potential realization. The nominal input sky map in conjunction with the fiducial CMB realization is then processed to include systematic effects such as detector pointing, and known point sources \cite{254}. From this point, independent realizations of noise are added to mimic the observational data stream \cite{255}. Finally, by subtracting out the fiducial CMB and foregrounds, the noise (+ systematics) simulations are obtained.

In addition to the $\text{DX12}$ data and $\text{FFP10}$ simulations, in this work we also make use of the
PR4 NPIPE data [220] and its corresponding simulations. At the level of the data, there are improvements in the polarization calibration procedure, and additional data (\(\sim 8\%\)) from the (previously neglected) satellite repointing periods, leading to a slightly reduced noise level. Enhanced data processing further accounts for systematic effects by building templates in the time-ordered data stream. These template amplitudes can be inferred via marginalization over the sky, and lead to iteratively “cleaned” maps as the template amplitudes converge to zero. With respect to the simulations, the lensed scalar CMB remains unchanged from FFP10, but the end-to-end simulations emulate many features from the data processing pipeline, such as the removal of systematics in time ordered information previously discussed. Finally, the NPIPE processing allows for the construction of maximally independent detector-set split maps, which are referred to as the A and B splits. These should in principle provide improved noise decorrelation with respect to one another than that between the half mission splits.

Despite the sophisticated instrumentation model used in FFP10 and NPIPE, small discrepancies between the power spectra of the simulations and those observed in the equivalent data will persist. Small spatially varying differences in power could bias our inference of primordial power modulation. Fortunately, the realization-dependent disconnected noise correction (\(N_{\text{RD},(0)}L\)) that we apply automatically corrects for the difference between the true and simulation covariance to leading order, making the estimators relatively robust. However, for higher-order differences to be small, we do still need to ensure that the power in the simulations is at least perturbatively close to the power in the data.

We therefore adjust the simulations so that the power spectra of the filtered data and filtered simulations match, targeting consistency to the sub-percent level. In order to do this, we compute the power difference from the WF data and the ensemble average spectrum from WF simulations: \(\Delta C_\ell \equiv C_\ell^{\text{data}} - \langle C_\ell^{\text{sims}} \rangle\). The corrections are Gaussian and isotropic realizations of power that are added to the simulations and/or data. We defer explicit details of this prescription to Appendix D where we explain how we simultaneously correct the auto and cross-spectra between datasets. We illustrate the effectiveness of this corrective procedure for the DX12 143/217 GHz data and simulations in Figure D.1.

### 3.3 Pipeline validation and calibration

We now validate our pipeline by simulating CMB skies modulated with a known \(\tau_{\text{NL}}\) modulation field before adding noise. We check that we recover the modulation spectrum correctly, and that the corresponding posterior is consistent with the input value. Since the reconstruction on a single realization is noisy, we check for biases over a set of simulations with the same input \(\tau_{\text{NL}}\). The mean of the reconstructed modulation power spectra should match the input spectrum, and the product of the posteriors should peak close to the input \(\tau_{\text{NL}}\).

We perform these tests using 300 FFP10 HM1/HM2 and NPIPE A/B simulations on the cut-sky,
using the same masks as in the final data analysis. In each simulation set, we use common scalar CMB simulations (from FFP10) modulated with realizations of the modulation field \( f(\hat{n}) \) with theoretical spectrum \( C^f_L = \tau_{NL} C^\zeta_L = 1500 C^\zeta_L \). Key differences in the simulation sets are found in their respective noise maps, as described in Section 3.2. After building the modulated simulations, we compute the WF and IVF fields, and then construct ensembles of modulation field estimators from map pairs as given in Eq. (3.25). In order to compute the mean field and \( N_L^{(0)} \) corrections we repeat this procedure for the same CMB + noise simulations without the modulation signal. Each modulation reconstruction is then mean field and \( N_L^{MC,(0)} \) subtracted. We use the same realizations of the noise and unmodulated lensed CMB for the with-signal and without-signal simulations, so that the Monte Carlo and cosmic variance noise largely cancels. This choice enables us to measure any required calibration factors out to small scales, even where the signal is well below the reconstruction noise.

With the ensemble of reconstructed spectra we compute the calibration factor by comparing their mean against the theoretical input spectrum: \( k_L = C^f_L^\text{Theory} / \langle C^f_L \rangle_{\text{sims}} \). Since the calibration is empirical (and therefore depends on the details of the simulations, mask and processing) it is not universally defined, and requires expensive calculation. We therefore only compute the calibration for the simulation sets discussed in this section; which include those relevant for our baseline results. In Figure 3.3 we plot the mean over reconstructed power spectra after \( k_L \)-correction, and in Figure 3.4 plot the product of the corresponding posterior distributions (normalized by their maximum value).

The \( k_L \) factors are plotted within the inset panels of Figure 3.3 for MV reconstructions using the FFP10 and NPIPE sets. We further find that the TT calibration factors are approximately equal to those given for MV. In order to compute \( k_L \) we first bin the raw empirical result. For the largest scales of interest \( L = 1, 2, 3, 4 \) we retain unit multipole bins, before incrementally expanding the bin width in multiples of three, weighting by an \( L(L+1) \) kernel. The binned solutions are then interpolated to build smooth functions over the smaller scales. This strategy is designed to avoid over-fitting fluctuations in the Monte Carlo noise, whilst modelling the large-scales accurately where the bulk of the
modulation signal resides. At first pass we found that the large-scale calibration leads to over-correction of the posterior estimates, and we therefore decrease the amplitude of $k_L$ by 5% on these scales.

In Figure 3.4 we plot the calibrated, uncalibrated and “idealized” posterior distributions. In the idealized case, we use the (known) realizations of input full-sky modulation power spectra in place of the reconstructions appearing in the estimates of (3.38). This serves as a second reference point in addition to the input $\tau_{NL} = 1500$ marker, with which the product of posteriors should be consistent. Notably this posterior does not peak at the precise fiducial value, which is driven by the cosmic variance and the limited sample size. After calibration, the product of reconstructed simulation posteriors are highly compatible with the product of the idealized posteriors, both of which agree with the theoretical value ($\tau_{NL} = 1500$) to within $1\sigma$. We have further checked that the calibrated posteriors of without-signal simulations are consistent with $\tau_{NL} = 0$ to the same degree.

![Figure 3.4: Product of posterior distributions (normalized to one) from 300 with-signal $\tau_{NL}$ simulation minimum variance reconstructions: Left: NPIPE (A, B); Right: FFP10 (HM1, HM2). Vertical dash-dotted lines indicates the input simulation value. Solid black posterior assumes perfect full-sky reconstruction of the input modulation power. Red and green lines indicate the uncalibrated and calibrated reconstructions respectively. All posterior distributions are constructed over the multipole range $L \in [1, 50]$.](image)

### 3.4 Results

We first report our results for the TT modulation reconstruction obtained from the 143 and 217 GHz channels, which are derived from DX12 data and FFP10 simulations. In Figure 3.5 we plot the uncalibrated power spectra for the auto- and joint-reconstructions between these data. In red we show the modulation field power spectrum after mean field and $N_{RD,0}^L$ subtraction, the latter of which is given in green for reference. In orange bands, we plot the 68%, 95% and 99% confidence intervals obtained directly from the equivalent modulation power spectra of without-signal simulations. This figure is directly comparable to that given in Figure 20 of the Planck 2013 non-Gaussianity constraints [217]. Beyond obvious differences in the data and simulations, we note that the methodology presented in this work is slightly different. In Planck 2013 the disconnected noise bias was computed from simulations alone ($N_{MC,0}^L$), which is not as robust as the realization-
dependent estimate \(N_{L}^{RD,(0)}\) used in this work. Moreover in this work we adjust the Gaussian simulation power measured against the data, which makes our \(N_{L}^{(0)}\) and mean field subtraction more optimal. When comparing Figure 3.5 to its equivalent in 2013, it should be noted that “(143, 217) GHz” in this work equates to “143×217 GHz” in the other. In Planck 2013 the 857 GHz intensity map was projected from the 143 and 217 GHz intensity maps whilst obtaining the IVF maps. This projection acted as a template for dusty foreground removal. We project 857 GHz in the same way but now using the DX12 data.

**Figure 3.5:** Uncalibrated modulation reconstruction power spectra from DX12 143 and 217 GHz data. From left-to-right we plot the auto 143 GHz, joint (143, 217) GHz and auto 217 GHz reconstructions. Red and green lines indicate the modulation and \(N_{L}^{RD,(0)}\) power spectra, and orange bands indicate the 68%, 95% and 99% C.L. of the simulation distribution with no signal.

With this in mind we note that there are differences between the reconstructed spectra, most vividly between the amplitude and sign of the largest scale reconstruction modes. In Planck 2013 there was an \(\sim 4\times\) dipolar power difference between the auto-reconstructions, with the power of the joint reconstruction roughly matching the 143 GHz auto. We find qualitatively similar results; though the difference is \(\sim 2\times\) and the sign of our 143 GHz dipole is now negative. In both analyses, it is likely that the 217 GHz auto signal is strongly contaminated by non-cosmological signal which is not present in the 143 GHz observations. The differences in the dipole sign are partially explained by the realization dependent noise modelling, which slightly boosts the amplitude of \(N_{L}^{(0)}\) in this work resulting in a more heavily subtracted signal. The octopole signal similarly has a frequency-dependent amplitude, which in the joint reconstruction approximately match the 143 GHz auto.

The most striking difference is in the quadrupole. Whilst we report that there is very little reconstruction signal in this mode across the auto- and joint-reconstructions, in Planck 2013 this dominated the autos before being heavily mitigated in the joint. The strong frequency-dependent amplitude seen in the Planck 2013 analysis already indicated that the bulk of the signal was not from primordial modulation. The difference we observe in this work is most likely explained by better quality (and power-corrected) simulations, which enable us to subtract the frequency-dependent components of the mean field more reliably. Though varying by a lesser extent, we observe a reduced amplitude in the 143 GHz auto, which could be explained in the same way.

We now consider the foreground cleaned SMICA maps, for which we compare DX12 and
NPIPE data. It is helpful to consider both of these data (and corresponding simulation) products since it enables us to identify possible systematic effects arising from the map-making procedure that could lead to a false detection. When comparing these data we perform modulation reconstruction for a variety of different sky fractions, the masks of which are shown in Figure 3.6. By analysing the response of the modulation reconstruction to lesser sky coverage, we check for possible galactic foreground contaminants. The baseline mask maintains \( f_{\text{sky}} \approx 67\% \), and is almost identical to that used in the Planck lensing constraints. Additional masks are then obtained by taking a threshold cut in the 353 GHz intensity emission (after smoothing the data to avoid removing regions of high-intensity CMB). We vary the threshold in order to obtain masks with an incremental 10% reduction in \( f_{\text{sky}} \).

Figure 3.6: Masks used to test foreground contamination. Black corresponds to the fiducial mask used in the Planck lensing analysis, coloured masks layered below are generated from a threshold cut in the smoothed intensity of emission in 353 GHz. By construction, the larger masks are a superset of the smaller ones, so that \( M_{37} \subseteq M_{47} \subseteq M_{57} \subseteq M_{67} \).

For NPIPE analyses we mask a small number of additional pixels, which correspond to the pre-processing mask, applied during our implementation of the SMICA map making process; this essentially leaves the sky fraction unchanged. For DX12 analyses (when we use the HM splits) the data contains missing pixels, therefore we apply further masking which results in a 1 – 3% reduction in \( f_{\text{sky}} \). In the following text, these masks will be referred to as M67, M57, M47 and M37, where any additional masking should be interpreted from the data it is applied to and the number convention indicates the sky fraction of the baseline masks in Figure 3.6.

Although NPIPE has more end-to-end simulations available than FFP10, to keep these initial comparisons on equal footing we construct mean field and \( \tilde{\bar{N}}_{L}^{(0)} \) corrections using the same number of simulations (80 for mean field, 220 for \( \tilde{\bar{N}}_{L}^{(0)} \)). In Figure 3.7 we present the results for \( \tilde{C}_{L}^{I} \) computed from minimum variance modulation reconstruction. In the top and bottom rows we plot the power spectra of joint reconstructions NPIPE (A, B) and DX12 (HM1, HM2) respectively, and left-to-right the reduced \( f_{\text{sky}} \). For a given baseline mask, NPIPE and DX12 have similar spectra, with the amplitude of the large-scale reconstruction modes of NPIPE being generically lower. All of the spectra have a shape consistent with the (143, 217) GHz reconstruction, in which most of the power is found in the octopole; which has decreasing power as a function of \( f_{\text{sky}} \). This systematic decrease in the octopole power suggests that there may be a galactic foreground contaminant in the modulation signal,
which is partially removed as the size of the galactic mask is increased. However, as we decrease the available sky fraction the statistical significance of the reconstructed signal decreases; as can be seen by the error bars expanding left-to-right. We do not present the TT reconstruction spectra since these are essentially the same as the MV, though the spectra and errors have a slightly larger amplitude.

In the upper part of Table 3.1 we provide estimates for the uncalibrated modulation power spectrum reconstructions. For each power spectrum we compute an uncalibrated $\bar{\tau}_{NL}$ estimate via direct application of Eq. (3.35), and provide a rough estimate of the error $\sigma(\bar{\tau}_{NL})$ from applying the same estimator to reconstructed modulation spectra of without-signal simulations and computing the resulting standard deviation. For each spectrum we form an estimate with $L_{\text{min}} = 1$ and consider $L_{\text{max}} = 2, 10$ and 50. We provide results for both TT and MV reconstructions.
Uncalibrated inverse-variance weighted estimates: $\bar{\tau}_\text{NL} \pm \sigma(\bar{\tau}_\text{NL})$ for $1 \leq L \leq L_{\text{max}}$

<table>
<thead>
<tr>
<th>Data</th>
<th>Mask</th>
<th>$L_{\text{max}} = 2$</th>
<th>$L_{\text{max}} = 10$</th>
<th>$L_{\text{max}} = 50$</th>
<th>$L_{\text{max}} = 2$</th>
<th>$L_{\text{max}} = 10$</th>
<th>$L_{\text{max}} = 50$</th>
</tr>
</thead>
<tbody>
<tr>
<td>DX12 (143, 217) GHz</td>
<td>M67</td>
<td>$-340 \pm 267$</td>
<td>$-56 \pm 272$</td>
<td>$-94 \pm 274$</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>DX12 (HM1, HM2)</td>
<td>M67</td>
<td>$-260 \pm 273$</td>
<td>$95 \pm 274$</td>
<td>$50 \pm 277$</td>
<td>$-230 \pm 269$</td>
<td>$113 \pm 267$</td>
<td>$49 \pm 269$</td>
</tr>
<tr>
<td>DX12 (HM1, HM2)</td>
<td>M57</td>
<td>$-459 \pm 301$</td>
<td>$-154 \pm 302$</td>
<td>$-185 \pm 306$</td>
<td>$-430 \pm 302$</td>
<td>$-145 \pm 294$</td>
<td>$-200 \pm 297$</td>
</tr>
<tr>
<td>DX12 (HM1, HM2)</td>
<td>M47</td>
<td>$-492 \pm 319$</td>
<td>$-177 \pm 330$</td>
<td>$-208 \pm 332$</td>
<td>$-500 \pm 323$</td>
<td>$-241 \pm 322$</td>
<td>$-291 \pm 323$</td>
</tr>
<tr>
<td>DX12 (HM1, HM2)</td>
<td>M37</td>
<td>$-384 \pm 346$</td>
<td>$-166 \pm 354$</td>
<td>$-207 \pm 357$</td>
<td>$-425 \pm 357$</td>
<td>$-302 \pm 357$</td>
<td>$-369 \pm 360$</td>
</tr>
<tr>
<td>NPIPE (A, B)</td>
<td>M67</td>
<td>$-46 \pm 260$</td>
<td>$129 \pm 259$</td>
<td>$79 \pm 261$</td>
<td>$1 \pm 248$</td>
<td>$167 \pm 248$</td>
<td>$107 \pm 248$</td>
</tr>
<tr>
<td>NPIPE (A, B)</td>
<td>M57</td>
<td>$-306 \pm 276$</td>
<td>$-145 \pm 274$</td>
<td>$-185 \pm 275$</td>
<td>$-288 \pm 267$</td>
<td>$-134 \pm 266$</td>
<td>$-186 \pm 266$</td>
</tr>
<tr>
<td>NPIPE (A, B)</td>
<td>M47</td>
<td>$-313 \pm 282$</td>
<td>$-107 \pm 286$</td>
<td>$-134 \pm 287$</td>
<td>$-290 \pm 273$</td>
<td>$-125 \pm 278$</td>
<td>$-165 \pm 279$</td>
</tr>
<tr>
<td>NPIPE (A, B)</td>
<td>M37</td>
<td>$-170 \pm 296$</td>
<td>$-66 \pm 299$</td>
<td>$-120 \pm 301$</td>
<td>$-130 \pm 285$</td>
<td>$-103 \pm 285$</td>
<td>$-174 \pm 285$</td>
</tr>
</tbody>
</table>

Calibrated inverse-variance weighted estimates: $\hat{\tau}_\text{NL}$, $\sigma(\hat{\tau}_\text{NL})$ for $1 \leq L \leq L_{\text{max}}$

<table>
<thead>
<tr>
<th>Data</th>
<th>Mask</th>
<th>$L_{\text{max}} = 2$</th>
<th>$L_{\text{max}} = 10$</th>
<th>$L_{\text{max}} = 50$</th>
<th>$L_{\text{max}} = 2$</th>
<th>$L_{\text{max}} = 10$</th>
<th>$L_{\text{max}} = 50$</th>
</tr>
</thead>
<tbody>
<tr>
<td>DX12 HM (HM1, HM2)</td>
<td>M67</td>
<td>$-321 \pm 326$</td>
<td>$12 \pm 317$</td>
<td>$-30 \pm 319$</td>
<td>$-287 \pm 321$</td>
<td>$39 \pm 309$</td>
<td>$-21 \pm 310$</td>
</tr>
<tr>
<td>NPIPE AB (A, B)</td>
<td>M67</td>
<td>$-62 \pm 290$</td>
<td>$86 \pm 279$</td>
<td>$36 \pm 280$</td>
<td>$-63 \pm 293$</td>
<td>$86 \pm 283$</td>
<td>$36 \pm 283$</td>
</tr>
</tbody>
</table>

Table 3.1: Summary of the inverse-variance weighted $\tau_{NL}$ estimates. For each combination of data, mask and modulation estimator, we compute $\bar{\tau}_{NL}$ estimates from $L = 1$ to $L_{\text{max}}$ using Eq. (3.35). Each entry in the table(s) corresponds to the data estimate and standard deviation of equivalent estimates from without-signal simulations. Upper: Uncalibrated estimates from DX12 and NPIPE data. Lower: Calibrated estimates.
When only \( L \in [1, 2] \) define the estimator we sometimes have large negative estimates, which are driven directly by the negative mode(s) in the reconstructed power spectra; these estimates are un-physical. For the remaining estimators when \( L_{\text{max}} \in [10, 50] \) we see that all of the data estimates are consistent with \( \tau_{\text{NL}} = 0 \). It should be noted that using the standard deviation of \( \bar{\tau}_{\text{NL}} \) from simulations is not the most rigorous approach to estimating the error, since the distribution is slightly skewed by the \( \chi^2 \) properties of the reconstruction modes. Our estimated error is however still useful as an approximate means of assessing the data consistency with \( \tau_{\text{NL}} = 0 \).

The error bars for the \( \bar{\tau}_{\text{NL}} \) constraints in Table 3.1 nearly saturate by \( L_{\text{max}} = 10 \), since the bulk of the signal is in the lowest multipoles. However, the measured values mostly shift downward moving from \( L_{\text{max}} = 10 \) to \( L_{\text{max}} = 50 \), by an amount that look anomalous given the little information that should be available in that range (hence the error bars being almost unchanged). This shift is dominated by the multipole range \( 10 \lesssim L \lesssim 40 \), corresponding to the run of mostly negative modulation power multipole estimates clearly seen by eye in e.g. Fig. 3.7. However, we find consistency with the simulations to \( \lesssim 3\sigma \) by applying Eq. (3.35) to each of the sub-intervals \( 5 + \delta L \leq L \leq 15 + \delta L \) for \( \delta L \in [0, 10, 20, 30] \). We have also found that \( \text{NPIPE} (A, B) \) is just over \( 3\sigma \) low for \( 9 \leq L \leq 35 \). While visually striking, negative signals cannot be produced by a physical modulation signal, and this multipole range is chosen a posteriori, so we do not regard this is significant evidence of a problem.

In addition to the data provided in the table, we report \( \bar{\tau}_{\text{NL}} \pm \sigma(\bar{\tau}_{\text{NL}}) = -150 \pm 579, 173 \pm 577 \) and \( 156 \pm 578 \) for \( L_{\text{min}} = 1 \) with \( L_{\text{max}} = 2, 10 \) and \( 50 \), from the TTxPP \( \text{NPIPE} (A, B) \) reconstruction spectrum using the M67 mask. As discussed in Section 3.2, estimators of this type have a large variance, and this can be seen directly by comparing the error on this result to the analogous case of MV \( \text{NPIPE} (A, B) \) given in Table 3.1, which is more than \( 2\times \) smaller.

In Figure 3.8 we show the corresponding posterior distribution of estimates for a range of \( L_{\text{max}} \) with fixed \( L_{\text{min}} = 1 \). All of the posterior distributions are consistent with \( \tau_{\text{NL}} = 0 \) with very long tails. In a few instances, the posterior peaks slightly away from zero, but in most cases peaks at zero. The dashed vertical lines in each panel indicate the upper bound on \( \tau_{\text{NL}} \) inferred from the 95\% C.L. on the \( L_{\text{max}} = 50 \) posterior distribution. This bound shifts as a function of \( f_{\text{sky}} \) but is relatively stable for the larger-sky fractions. Note that these posteriors have no calibration factors of \( k_L \), but do contain the Monte Carlo and Hartlap corrections to their covariances as described in Section 3.1.4. Without applying the full calibration these estimates are slightly too small.

All of the uncalibrated posterior constraints are provided in the upper portion of Table 3.2. We provide the constraints from using data up to \( L_{\text{max}} = 2, 10, 50 \) in both temperature and polarization (except for \( \text{DX12} \ (143, 217) \) GHz). In general, the upper-bound decreases as more multipoles are included in the likelihood analysis; with the tightest constraints arising when we preserve the larger sky-fractions. This latter trend is associated with
the increased (co-)variance of modulation estimators as the sky fraction is decreased, leading to broader posteriors. For M67 and M57, we find that the 95% upper bound lies approximately between 1500-2000.
Uncalibrated $\tau_{NL}$ constraints: 95\% C.L. on $P(\tau_{NL}|\bar{\tau}_{NL})$ for $1 \leq L \leq L_{\text{max}}$

<table>
<thead>
<tr>
<th>Data</th>
<th>Mask</th>
<th>Temperature</th>
<th>Minimum variance</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$L_{\text{max}} = 2$</td>
<td>$L_{\text{max}} = 10$</td>
<td>$L_{\text{max}} = 50$</td>
</tr>
<tr>
<td>DX12 (143, 217) GHz</td>
<td>M67</td>
<td>3300</td>
<td>2600</td>
<td>2000</td>
</tr>
<tr>
<td>DX12 (HM1, HM2)</td>
<td>M67</td>
<td>3400</td>
<td>3100</td>
<td>2200</td>
</tr>
<tr>
<td>DX12 (HM1, HM2)</td>
<td>M57</td>
<td>4100</td>
<td>2300</td>
<td>1800</td>
</tr>
<tr>
<td>DX12 (HM1, HM2)</td>
<td>M47</td>
<td>5200</td>
<td>4100</td>
<td>3100</td>
</tr>
<tr>
<td>DX12 (HM1, HM2)</td>
<td>M37</td>
<td>8700</td>
<td>3900</td>
<td>3600</td>
</tr>
<tr>
<td>NPIPE (A, B)</td>
<td>M67</td>
<td>3600</td>
<td>2200</td>
<td>2000</td>
</tr>
<tr>
<td>NPIPE (A, B)</td>
<td>M57</td>
<td>4000</td>
<td>2200</td>
<td>1600</td>
</tr>
<tr>
<td>NPIPE (A, B)</td>
<td>M47</td>
<td>3500</td>
<td>3300</td>
<td>2600</td>
</tr>
<tr>
<td>NPIPE (A, B)</td>
<td>M37</td>
<td>6200</td>
<td>2600</td>
<td>2200</td>
</tr>
</tbody>
</table>

Calibrated $\tau_{NL}$ constraints: 95\% C.L. on $P(\tau_{NL}|\hat{\tau}_{NL})$ for $1 \leq L \leq L_{\text{max}}$

<table>
<thead>
<tr>
<th>Data</th>
<th>Mask</th>
<th>Temperature</th>
<th>Minimum variance</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$L_{\text{max}} = 2$</td>
<td>$L_{\text{max}} = 10$</td>
<td>$L_{\text{max}} = 50$</td>
</tr>
<tr>
<td>DX12 (HM1, HM2)</td>
<td>M67</td>
<td>16000</td>
<td>3900</td>
<td>2800</td>
</tr>
<tr>
<td>NPIPE (A, B)</td>
<td>M67</td>
<td>5900</td>
<td>2200</td>
<td>1900</td>
</tr>
</tbody>
</table>

Table 3.2: Summary of upper limits on $\tau_{NL}$ from posterior probabilities. **Upper:** Uncalibrated data constraints from DX12 and NPIPE data; **Sep.** refers to the separate frequency constraint from the (143, 217) GHz cross-estimator. **Lower:** 95\% C.L. $\tau_{NL}$ constraints from calibrated analyses. The NPIPE (A, B) constraint in the lower table is our baseline result.
We finally produce results for the calibrated analyses of TT and MV modulation reconstruction using DX12 (HM1, HM2) and NPIPE (A, B) map pairs. For the calibrated NPIPE analysis, we have now used all available simulations; 120 and 480 simulations for the mean field and noise bias estimates respectively. In the lower part of Table 3.1 we provide the inverse-variance weighted estimates for these data. For the $L_{\text{max}} = 2$ constraint, we see that the estimate is pulled more negative compared to the uncalibrated result. This is due to the dipole and quadrupole receiving multiplicative corrections with $k_{L} > 1$ (see inset on Figure 3.3) and the fact that the reconstructed spectra are negative for these two modes. Once further multipoles are included, the estimates become more consistent with zero. For all of the calibrated analyses the error estimate becomes slightly larger; which reflects the high weighting that the low multipoles receive for this particular form of the estimator and, as previously noted, that $k_{L} > 1$ on these scales.

In Table 3.2 we present the calibrated $\tau_{NL}$ constraints on the posterior probability, which exhibit relatively large changes from their respective uncalibrated constraints. This is especially notable for the DX12 95% C.L. constraints for $L_{\text{max}} = 2$. For these calibrated results, the posterior peaks at zero, but has a long and shallow tail, which leads to the relatively weak constraining power. As in the uncalibrated analyses, all of the posterior distributions remain consistent with zero. After calibrating, however, the posterior peaks are pulled slightly positive, such that the normalized distributions exhibit greater shifts as a function of $L_{\text{max}}$. For our baseline result we advocate the calibrated MV NPIPE (A, B) analysis, for which we report $\tau_{NL} < 1700$ 95% C.L. The full-calibrated result for the modulation power spectrum and the corresponding posterior distributions are shown in Figure 3.9.

Finally, we attempt to gauge the impact improved data quality has on $\sigma(\tau_{NL})$ estimates using a semi-analytic approximation of the reconstruction noise bias $N_{L}^{(0)}$. These $N_{L}^{(0)}$ can be computed cheaply from the power spectra of the IVF maps of different data, and we do so for the NPIPE, DX12 and Planck 2013 (PR1) data. Using these spectra we simulate $10^6$ realizations of the Gaussian noise bias following a $\chi^2$ distribution, and propagate the spectra into Eq. (3.35) to build $\bar{\tau}_{NL}$ estimates. The standard deviation
from these ensembles is then used to build approximate errors for each dataset, and by further splitting the estimates into \(10^3\) subsets before computing \(\sigma(\tau_{\text{NL}})\) we estimate the sample-variance from the scatter on the errors. The results from this exercise are given in Figure 3.10 for \(L_{\text{min}} = 1\) and \(L_{\text{max}} = 10\), and we see that these are highly consistent with those computed in Table 3.1 for the full end-to-end simulations.

Moving from PR1 to DX12 for (143, 217) GHz analyses, there is a large improvement from the overall reduction in noise. However when moving from TT to MV (relevant for NPIPE (A, B) and DX12 (HM1, HM2)) there is a relatively small improvement, which highlights that at Planck polarization noise levels the reconstruction signal is largely accounted for by the temperature modes. The forecast suggests that there is \(\sim 20\%\) reduction on the error (assuming a null hypothesis) when using NPIPE rather than PR1 data. Interestingly the errors for DX12 (143, 217) GHz seem to outperform those from DX12 (HM1, HM2). In the analytic model, this is explained by the competing factors of noise levels and sky fraction: Although the HM maps have slightly lower noise properties, they require additional masking due to missing pixel data, and this few percent difference in \(f_{\text{sky}}\) propagates directly into the error estimate. It should further be noted that the component-separated maps have a better control over galactic and astrophysical contaminants, the details of which are not folded into this analytic calculation. The component-separated maps are therefore expected to produce cleaner estimates of primordial modulation despite the seemingly competitive reconstruction noise bias of the separate frequency maps.

![Figure 3.10: Semi-analytic error forecasts for \(\tau_{\text{NL}}\) estimates with different data choices: Diamond markers correspond to MV and the remaining circular markers to TT estimators.](image)

### 3.5 Conclusions

Using the latest Planck data, we found the significantly tightened constraint \(\tau_{\text{NL}} < 1700\) 95% C.L., reducing the upper limit by a factor of \(\sim 1.7\) compared to previous results. A Universe with purely Gaussian primordial fluctuations remains consistent with the data, though the upper limit remains substantially larger than would be predicted in most multi-field inflation models.
In addition to making use of new data scans and polarization, and using the improved NPIPE processing, we also made several significant improvements to the analysis methodology, including a more optimal and robust model for the connected biases and use of foreground-separated maps. The quadratic estimator pipeline is very similar to that for CMB lensing, but with the additional requirement to consistently model the expected lensing four-point signal. Several lessons from this work have already been directly useful to inform the analogous updated NPIPE CMB lensing analysis [256]. The constraint is still dominated by information from the temperature due to the relatively low signal to noise in the Planck polarization, but could be significantly improved in future with small-scale low-noise polarization data. More dramatic improvements may be possible in conjunction with low shot noise large-scale structure surveys and small-scale kinetic SZ observations [257].

The current analysis also could be further optimized by more optimal filtering and handling of the noise anisotropy (e.g. see Ref. [258] in the context of lensing), but given the signal is temperature dominated, the constraint is not expected to improve dramatically. Future work could also easily extend the analysis to scale-dependent $\tau_{NL}$-like signals.
Chapter 4

Closing remarks

Inflationary cosmology continues to succeed in providing a general description for the origin of the initial curvature perturbations. As we have recapitulated across the introductory background material of this thesis, these initial perturbations can explain the growth of inhomogeneities that describe the statistical properties of the microwave background, and ultimately the later large-scale structure that we observe today. This is more than a qualitative success, indeed, theoretical methods are now mature enough that we can make highly predictive statements based on the functional description of an inflationary model, in terms of its potential and field space geometry. Performing such predictions exceed the level of the power spectrum, and we can readily compute non-Gaussian statistics in the form of the bispectrum; which will continue to serve as a key identifier for multiple field as information is scoured from the observational fields of next generation cosmology surveys.

In this thesis, we have pushed to understand inflationary cosmology from multiple sides, we have:

1. developed automated sampling software to explore models using the Transport method;
2. deployed such software to investigate the famously delicate model of $D3$-brane inflation, and explored its corresponding bispectrum statistics; and
3. derived a novel constraint on the amplitude of the primordial trispectrum $\tau_{NL} < 1700$ (95% C.L.), universally lowering the ceiling on permitted multiple field inflationary models.

A clear vein of this thesis has been the connection between inflationary cosmology and ultraviolet physics; specifically in connection to the $D3$-brane scenario. It has been demonstrated that the $D3$-brane model has undesirable spectral properties at the level of the power spectrum. However, whilst this statement is deterring at the ‘ensemble’ level, there are certainly realisations of the model which are compatible with (the likelihood from) Planck. Without advocating anthropic arguments on which to assess the need for the model to generically predict the cosmology we see today, we note that there are identifiable areas in which we could improve the models generic compatibility with data.
An appropriate ‘adjustment’ to the scalar D3-brane potential would be one that flattens the inflexion point for a longer duration, in a way that promotes observationally relevant modes exiting the horizon about this point. Small and weakly radial dependent corrections could achieve this, though (as emphasised in [89]) such corrections cannot be added ‘by hand’ in string theory, and must manifest from fundamental changes in the stringy background data. This is a key reason why realising inflation from string-theory is so hard to achieve concretely, and why the D3-brane scenario is somewhat remarkable in its own right. Notwithstanding this obstacle, some authors have commented on configuration changes that could effect the radial profile of the potential. One such example is the descent of the D7-branes wrapping the four-cycle into the warped throat region [146]. This would adjust the radial profile, though would lead to wider - possibly undesirable - changes in other sectors of the potential. It is anticipated that these radial changes are ‘theoretically’ possible, though non-trivial to develop in practice.

It would be interesting to compute the statistical properties of the D3-brane trispectrum, and in particular, to forecast what is to be expected for $\tau_{\text{NL}}$. An extension of the Transport method has readily been presented in the literature for multiple field models carrying a Euclidean field space metric [111]. These tools are yet to be packaged into the general purpose software PyTransport or CppTransport, though this development is likely to occur in the future.

One such challenge to anticipate in this development is combinatorics: for each Fourier mode there are $2N_f(N_f + 1/2)$ and and $8N_f^3$ equations to be solved for the 2- and 3-point correlation function of field perturbations [75]. Neglecting symmetries, the trispectrum will scale to require roughly $\sim N_f^4$ equations to be solved. Secondly, the trispectrum will require the 4th order covariant derivative of the potential to be computed. Whilst this remains straightforward to achieve via computer algebra, it is worth emphasising that the structure of the D3-brane potential contains thousands of non-separable terms (for $\Delta < 3.8$), as such, this additional differentiation will generate an appreciable number of new terms. This additional complexity will lead to a large volume of source code for the D3-brane model (and those of similar complexity). Ultimately it is the compilation of these sources into a numerical program that could prove to be computationally prohibitive, particularly when using the typical PC hardware of today. Note that for reference, it currently takes $\mathcal{O}(1 \text{ hour})$ to compile the D3-brane model (including 3-point correlations). It is inevitable, however, that there will be continual advances in compiler technology and CPU memory such that this concern will be - at least partially - alleviated.

For these reasons, the software developments are beyond the immediate scope of this thesis to perform a detailed trispectrum analysis. Instead, we make use of the $\delta N$ formalism to obtain a rough estimate of $\tau_{\text{NL}}$ in the $D3$-brane model. This calculation only requires knowledge of the background data, from which, gradients of the expansion history can be computed numerically.

\footnote{The authors in fact comment on the D3-brane model discussed in this thesis! However, their numerical evaluation adopts a simplified version of the model which should not be compared to that presented in this work.}
To aid this discussion we only consider two ‘interesting’ samples: ‘Sample A’, which satisfies the Planck constraints on $n_s$; and ‘Sample B’, which generates a sizeable non-Gaussianity of $f_{NL}^{eq} \sim -5$ (as computed with the tree-level transport method). Implementing Eqs. (1.98, 1.100) we estimate the non-linearity parameters for the two samples, as shown in Fig. 4.1. For each sample, we plot $f_{NL}$, $\tau_{NL}$, and the single-field relation $\tau_{NL} = (6 f_{NL}/5)^2$. We also show the horizon exit time, $N_{Exit}$; relevant for Planck CMB scales under the assumption of instantaneous reheating.

Firstly, we note that we do not achieve the modestly large bispectrum associated with the Transport calculation for Sample B. This mismatch is not particularly surprising, since the $\delta N$ approximation we have deployed does not account for sub-horizon evolution in the field perturbations, and assumes that they are Gaussian distributed at horizon exit.

Sub-horizon correlations are captured within the transport method, however, and correlations between field perturbations on sub-horizon scales can persist ultimately to the late-time correlations of $\zeta$. In Fig. 4.2 we plot the background evolution for the corresponding samples. Subsequent correlations in the field perturbations are driven by the abrupt transitions in the angular fields.\(^2\) Inspecting the right hand panel (for Sample B)

\(^2\)Note that this observation is discussed in more detail in Sec. 2.3.2.4 for a different sample of interest.
we see that these transitions occur at about the same time as when CMB mods exit the horizon, and hence this effect is likely responsible for the computation of $|f_{NL}| \sim 5$.

From our $\delta N$ calculation, we see that there are clear indicators of multi-field effects. These can be diagnosed with transient excursions where $\tau_{NL} > (6f_{NL}/5)^2$. Alternatively, these instances can be visualised in Fig. E.1, where we plot the ratio of the left- and right-hand terms of this inequality. Interestingly, over the initial trajectory (corresponding roughly to the subhorizon regime with respect to the scale tracked in the Transport system), we indeed observe multiple field effects. After this time, the sample evolves in a single field type way. This aligns with our earlier discussion on D3-brane inflation in Chpt. 2, whereby we report many trajectories falling onto an effective single-field trajectory.

Based on these observations, a reasonable expectation is that $\tau_{NL}$ will be generally small (when the trajectory is effectively single field), though is likely to occasionally exhibit instances of $\tau_{NL} \sim \mathcal{O}(1000)$ when multiple field effects are large; somewhat consistent with the infrequent $f_{NL} \sim \mathcal{O}(100)$ in the greater ensemble.

In the future, CMB constraints on $\tau_{NL}$ via the modulation formalism will improve by virtue of greater signal to noise, and signal measurements out to smaller scales. An ideal experiment would be able to measure the ‘full sky’, as to reduce the variance in the large-scale modulating modes we are seeking to reconstruct.

Ultimately these ingredients collectively fold into the reconstruction noise power spectrum, which fundamentally limits the accuracy to which estimates can be formed from quadratic estimators. Taking the full sky, and assuming an experimental configuration otherwise crudely resembling *Planck*, in Fig. 4.3 we speculate the error we could expect on a reconstructed $\tau_{NL}$ power spectrum for a range of experimental setups. These estimators take the form of Eq. (3.34), with $(L_{\min}, L_{\max}) = (1, 50)$. Moreover, we assume an analytic Gaussian reconstruction noise (of MV flavour) in place of the reconstruction spectrum. Assuming a Gaussian beam of 5 arcmin, the required CMB noise spectra can be generated from scalar noise values $N^T_{\text{Level}}$ and $N^P_{\text{Level}}$. Note that for simplicity we have assumed that $N^P_{\text{Level}} = \sqrt{2}N^T_{\text{Level}}$. 

**Figure 4.2:** Background evolution of the scalar field coordinates associated with the D3-model for Sample A (left) and Sample B (right).
Figure 4.3: Estimated error on futuristic $\tau_{\text{NL}}$ power spectra, assuming a null hypothesis. Note that $L_{\text{max}}^{\text{CMB}}$ refers to the maximum multipole measured out to by the theoretical CMB experiment (or at least that which is retained after filtering).

In Fig. 4.3 it can be seen that reducing the noise reduces the error forecast, which is straightforwardly interpreted as a measurement of cleaner signal. There is a degeneracy between $L_{\text{max}}^{\text{CMB}}$ and $N_{\text{Level}}^{T}$ which corresponds to a trade off between ‘quality’ and ‘quantity’ of the CMB modes measured, though increasing $L_{\text{max}}^{\text{CMB}}$ ultimately leads to diminishing returns. The precise nature of this latter relationship will depend on the experimental configuration at hand; where lower noise translates to greater accessibility to small scale fluctuations.

In order to assess prospective future constraints, we use the likelihood model discussed in Chpt. 3, and consider a highly idealised experiment which can measure $L_{\text{max}}^{\text{CMB}} = 4000$ and $N_{\text{Level}}^{T} = 5.0$, as well as a slightly more realistic one with $L_{\text{max}}^{\text{CMB}} = 3000$ and $N_{\text{Level}}^{T} = 12.5$. In each configuration we choose $N_{\text{level}}^{P} = \sqrt{2}N_{\text{Level}}^{T}$ as before. These configurations are referred to as ‘Config. A’ and ‘Config B’ respectively. In Fig. 4.4 we model the covariance analytically, and assume that $C_{L}^{\tau_{\text{NL}}}=C_{L}^{\tau_{\text{NL}}}$ is measured perfectly, and plot the corresponding posterior distribution of estimates for $L \in [1, 50]$. In the left and right-hand panels we assume $\tau_{\text{NL}} = 50$ and $\tau_{\text{NL}} = 1000$ respectively, and mark the corresponding 95 % C.L. with a dashed-lines; which are obtained by integrating the posterior distribution from bottom-up.

From this figure we see that for $\tau_{\text{NL}} = 50$ it would not be possible to claim a detection, despite the low noise and high $L_{\text{max}}^{\text{CMB}}$ experiment. There would however be a significant constraint of $\tau_{\text{NL}} \lesssim 700$ across both configurations. By comparison, $\tau_{\text{NL}} = 1000$ generates a 2-tailed distribution for each experimental setup, clearly ruling out $\tau_{\text{NL}} = 0$ at greater than 3$\sigma$. Indeed, one might claim a statistical detection of $\tau_{\text{NL}}$. It should be borne in mind, however that this idealised analysis omits a number of key elements pertaining to a ‘realistic’ analysis. We comment that: 1. a Gaussian realisation of the modulation field is highly-unlikely to produce a power spectrum that perfectly traces $C_{L}^{\tau_{\text{NL}}}$ (the consequence of ‘fluctuations’ in this sense would pull the posterior adrift); 2. There are
**Figure 4.4:** Approximate posterior distribution of estimates for $\tau_{NL}$ assuming a perfect reconstruction and Gaussian noise on the full sky. *Left:* $\tau_{NL} = 50$; *Right:* $\tau_{NL} = 1000$.

Intrinsic fluctuations in not only the Gaussian noise field (but within the cosmological CMB itself) that can mimic a ‘modulation’ effect; 3. There are real-space anisotropic noise and astrophysical contaminants that could be confused with a modulation signal. The second of these two points is quite subtle, but it is a fundamental source of uncertainty when discriminating $\tau_{NL}$ from a cosmological background even under idealised conditions.

In the near future it will be exciting to see how constraints on primordial non-Gaussianity will improve with the advent of high quality large-scale structure (LSS) data. Whilst there are certainly more modes measurable from 3D volumetric surveys, their cross-correlation with high quality CMB data will likely serve instrumental in understanding primordial non-Gaussianities. The rationale behind this statement is as follows. Firstly, CMB theoretical predictions are highly mature and well understood, largely in terms of linear perturbation theory. This contrasts the more complicated non-linear modelling required to understand LSS formation. Secondly, the CMB secondary signals carry intrinsic correlations with LSS, and as such, can be used to discriminate non-Gaussian initial conditions via a ‘non-linear bias’. Indeed, there have been numerous articles over the years that explore these ideas (see e.g. Refs. [259, 260, 261, 262, 263]). Recently, a method for constraining $\tau_{NL}$ using the kinetic Sunyaev-Zel’dovich effect was presented in Ref. [264], suggesting that the amplitude of $f_{NL}$ and $\tau_{NL}$ may be measured with some $\sigma(f_{NL}) = 0.59$ and $\sigma(\tau_{NL}) = 1.5$, assuming optimistic CMB-S4 [265] and VRO [266] (formerly LSST) sensitivities. In terms of our optimistic error of $\tau_{NL}$ from an idealised experiment, this represents a remarkable $\sim 50 \times$ improvement.

**The future is bright.**
Appendix A

Minimal example: double quadratic model

We briefly demonstrate the installation process of the PyTransport Sampler. The user can immediately install the full software stack as follows with git:

```
git clone https://github.com/bkmarzouk/pytsa.git
cd pytsa
python -m pip install -e .
```

Note that for the installation to be successful, it is assumed that the user is running a Linux system, with a version of Python $\geq 3.8$, and GCC\(^1\), the GNU compiler collection.

Assuming the installation has been successful, let us define the double quadratic model

$$V(f_1,f_2) = \frac{1}{2}m_1^2 f_1^2 + \frac{1}{2}m_2^2 f_2^2,$$

equipped with the field space metric

$$G_{IJ} = \begin{pmatrix} R^2 & 0 \\ 0 & R^2 \sin^2 f_1 \end{pmatrix}.$$

In order to install this model, we must define a Python script that translates symbolic expression into C++ and compiles the result as an extended Python module.

```
import sympy as sym
from pytsa.pytrans_setup import Translator

# Number of fields
nF = 2

# Number of params (potential & metric)
nP = 3

# Build symbolic array for fields
f = sym.symarray('f', nF)
```

\(^1\)https://gcc.gnu.org/
# Build symbolic array for params
p = sym.symarray('p', nP)

# Construct symbolic expression for potential
V = sum([sym.Rational(1, 2) * f[i] ** 2 * p[i] ** 2 for i in range(2)])

# Construct symbolic expression for metric
G = sym.Matrix([[p[2] ** 2, 0], [0, p[2] ** 2 * sym.sin(f[0]) ** 2]])

# Translate model into c++ source code
Translator(nF, nP, V, G=G)

# Compile and install the submodule
Translator.install('dquad_2sphere')

This module supports the functionality described in the PyTransport 2.0 documentation by Mulryne & Ronayne [80], with additional functionality that allows the user to compute slow-roll parameters ($\epsilon_V$ and $\eta_V$) in addition to the mass-matrix $M_{IJ}$, which are retrievable via the compiled functions model.Epsilon, model.Eta and model.massMatrix. The relevant arguments for these functions and their description can be obtained through the built-in Python help function, e.g. help(model.Epsilon), which is now supported for all of the PyTransport compiled functionality.

To build a sampler, one must define the prior probabilities for parameters and initial conditions for field space, which will in turn be sampled over when the program is deployed.

```python
from pytsa.models import dquad_2sphere
from pytsa.sampler import setup_sampler
from scipy.stats import uniform, chi2, expon

# Example sampler setup file for double quadratic model.

# Initialize sampler object with the model (imported at the tope) and a name
# (for identification)
setup = setup_sampler.SamplerMethods(dquad_2sphere, "dquad_ex")

# Setup analysis hyper parameters
setup.set_analysis_params(tols=[1e-5, 1e-5])

# Setup field initial conditions; here we sample on the domain [-20, 20]
setup.set_field(0, 1, method=uniform(-20, 40))

# Setup dot field initial conditions; here we just use slow roll values
setup.set_dot_field(0, 1, method="sr")

# Setup parameter sampling; here we sample each mass parameter from an
minimal example: double quadratic model

exponential distribution
setup.set_param(0, 1, method=expon(scale=0.01))

# We also sample the metric parameter, R. Here we sample from the chi2
distribution.
# We chose this dist. since it is positive definite and the metric depends on
# R^2, so we don't over sample.
setup.set_param(2, method=chi2(3))

# Finally, build the sampler
setup.build_sampler()

By default, the sampler will be built and located within a subdirectory to the repository
root: pytsa/samplers.

The sampler program is then deployable via commands of the form

```
mpiexec -n 48 python pytsa/samplers/dquad_ex/run.py example_run --n_samples 10000 --latin --ns --eq
```

in a way that is readily suitable for local hardware or cluster systems. In this specific
event, the program is distributed over MPI using 48 cores via `mpiexec -n 48`. The
Python program is initialised with `python pytsa/samplers/dquad_ex/run.py`. 10,000
samples are requested with the flag and argument combination `--n_samples 10000`. A
Latin hypercube arrangement for the sampling space is defined with the flag `--latin`.
A calculation for the spectral index, \( n_s \) is requested with `--ns`. A calculation for the
equilateral reduced bispectrum \( f_{NL}^{eq} \) is requested with `--eq`.

A comprehensive list of available options can be obtained by requesting ‘help’ via:
`python path/to/sampler/dir/run.py --help` which will produce the following terminal
output for inspection

Configure PyTransport sampler Routine

positional arguments:
  name Name for sampling routine: defines subdir

optional arguments:
  -h, --help show this help message and exit
  --verbose Run sampler in verbose mode
  --apriori If flagged, then run sampler in apriori mode
  --latin If flagged, then run sampler in latin hypercube mode
  --n_samples N_SAMPLES Number of samples to compute
  --entropy ENTROPY System entropy value: Enables reproducible results
  --ns If flagged, computes 2-point data
--eq
  If flagged, computes 3-point data in equilateral mode

--fo
  If flagged, computes 3-point data in folded mode

--sq
  If flagged, computes 3-point data in squeezed mode

--alpha ALPHA [ALPHA ...]
  Pass values of alpha (Fergusson Shellard convention) to compute custom fnls. NOTE: Requires equal number of beta definitions.

--beta BETA [BETA ...]
  Pass values of beta (Fergusson Shellard convention) to compute custom fnls. NOTE: Requires equal number of alpha definitions.

Whilst this arrangement is highly specific, the user is free to explore any other configuration suitable to their analysis. The core outputs from a sampling job come in two forms: text files which can be directly read into GetDist [267] for comprehensive sample analysis, and a Pandas [268] data frame to support flexible analytics on the user end.
Appendix B

Transport equations for the spectral index

In this Appendix we briefly explain the transport computation of the scalar and tensor spectral indices. The superhorizon version of this calculation was given in Ref. [110], and extended to subhorizon scales in Ref. [79]. The discussion given here follows Ref. [79] with some refinements.

B.1 Scalar spectral index

First consider computation of the spectral index for \( \zeta \). As explained in §2.2.1.2, we start from the two-point function (2.49) on phase space, and define the ‘spectral matrix’ \( n^{AB} \) to satisfy [110]

\[
n^{AB}(k) = \frac{d\Sigma^{AB}(k)}{d\ln k} .
\]  

(B.1)

The \( \zeta \) two point function (2.43a) can be expressed at tree level in terms of \( \Sigma^{AB} \) using

\[
P_{\zeta} = N_{A} N_{B} \Sigma^{AB} + O(1 \text{ loop}),
\]  

(B.2)

where the coefficients \( N_{A} \) may be obtained from standard perturbation theory or the separate universe formula, and are momentum-independent when the mode \( k \) is more than a few e-folds outside the horizon [72, 127]. It follows that

\[
\frac{dP_{\zeta}}{d\ln k} \approx N_{A} N_{B} n^{AB} ,
\]  

(B.3)

from which the result (2.50) for \( n_s \) can be obtained.

Transport equation.—A transport equation for \( n^{AB} \) can be obtained by differentiating the transport equation for \( \Sigma^{AB} \). This yields

\[
\frac{dn^{AB}}{dN} = u^{A}_{C} n^{CB} + u^{B}_{C} n^{AC} + \frac{du^{A}_{C}}{d\ln k} \Sigma^{CB} + \frac{du^{B}_{C}}{d\ln k} \Sigma^{AC} + O(1 \text{ loop}).
\]  

(B.4)

The result is a coupled system of ordinary differential equations for \( \Sigma^{AB} \) and \( n^{AB} \). For the definition of the ‘\( u \)-tensor’ \( u^{A}_{B} \), see Dias et al. [79, 75]. When the transport equation (B.4)
Transport equations for the spectral index

B.1

is expressed with $N$ as the independent variable, it can be written

$$u^A_B = \begin{pmatrix} 0 & \delta^A_B \\ \mathcal{M}^A_B & -(3-\epsilon)\delta^A_B \end{pmatrix},$$

where $A$, $B$ are species labels corresponding to the phase space indices $\mathcal{A}$, $\mathcal{B}$. Recall that these indices run over the fields and momenta. Therefore, when a phase space index such as $A$ runs over its entire range the corresponding species label $A$ runs over its range twice. The matrix $\mathcal{M}^A_B$ is an effective mass matrix,

$$\mathcal{M}^A_B = -\frac{k^2}{a^2 H^2} \delta^A_B - \frac{M^A_B}{H^2},$$

where $M^A_B$ is the species mass matrix (2.48) appearing in the Lagrangian. Eq. (B.6) shows that the $k$-dependence of $u^A_B$ is exponentially suppressed on superhorizon scales. However, in this case it cannot be neglected. In the formula $P_\zeta \approx N^A_B \Sigma^{AB}$ we evaluate $N^A_B$ only far outside the horizon where $k^2/(a^2 H^2)$ is entirely negligible. In (B.4) we integrate over times when $k$ is comparable to or smaller than the horizon scale. For this regime the $k^2/(a^2 H^2)$ term is important in fixing the correct amplitude of $n^{AB}$ at horizon exit and cannot be dropped.

Initial conditions.—Eq. (B.4) can be solved simultaneously with the transport equation for $\Sigma^{AB}$, and future versions of CppTransport will do this by default. (The necessary functionality is already in the 201901 branch of the GitHub repository for those who wish to make use of it.) To proceed we require suitable initial conditions. A suitable set of initial values were given by Dias et al. [79], but these were truncated at leading order. In this paper, in an attempt to improve convergence with fewer e-folds of subhorizon evolution, we include subleading terms in both the slow-roll expansion and $k/(aH)$.

A next-order expression for the scalar two-point function was given in Ref. [202],

$$\langle \delta X^A(k_1,t_1) \delta X^B(k_2,t_2) \rangle = (2\pi)^3 \delta(k_1 + k_2) \left[ w(t_1)^A \right]^\dagger w(t_2)^B,$$

where the superscript ‘$\dagger$’ denotes complex conjugation, $A$ is an index in the field tangent space at time $t_1$, $B$ is a similar index at time $t_2$, and indices $I$, $J$, ..., label the field tangent space at an arbitrary earlier time $t_* < t_1, t_2$ characterized by horizon exit of a fiducial mode $k_*$. As usual, all of these indices should be raised and lowered with the field-space metric evaluated at the appropriate field-space coordinates. We are interested in the equal-time commutator for which $t_1 = t_2 = t$, with $t$ (corresponding to conformal time $\eta$) not too far from horizon exit of the mode $k = |k_1| = |k_2|$. To avoid the logarithm $\ln(-k_*\eta)$ in Eq. (B.8) below becoming large at this time we must usually choose $t_*$ to be roughly comparable to $t$. The usual choice is to take $t_*$ to be the moment when $k/(aH) = 1$. Up to next-order in the slow-roll expansion, but without making any approximation for the

---

1In Ref. [79] the initial conditions are given for a different matrix $\tilde{n}^{AB} = (d/d\ln k)(k^3 \Sigma^{AB})$, in which the factor $1/k^3$ in the ordinary power spectrum is removed. With this definition, Eq. (2.50) should be adjusted so the numerical constant 4 appearing on the right-hand side becomes 1.
explicit time dependence, the wavefunction matrix $w^A_I$ can be written

$$w(t)^A_I \equiv \frac{i}{\sqrt{2k^3}} H_\epsilon \Pi^A_K \left( (1 - ik\eta)e^{ik\eta}\delta^K_I + \epsilon_4 \left[ \ln(-k_\epsilon\eta) - 1 \right](1 - ik\eta)e^{ik\eta}\delta^K_I \right.$$

$$+ \left( \epsilon_6^\epsilon - \frac{M^K_I}{3H^2} \right) \left[ 2e^{ik\eta} + i\frac{\pi}{2}(1 - ik\eta)e^{ik\eta} - (1 + ik\eta)e^{-ik\eta} \int_{-\infty}^{k_\eta} \frac{dz}{z} e^{2iz} \right] \right)$$

(B.8)

where a subscript ‘$\epsilon$’ indicates evaluation at the time $t_\epsilon$, $\eta \approx \int_{t_\epsilon}^{\infty} dt'/a(t')$ is the conformal time corresponding to the evaluation time $t$ of the wavefunction, and $\Pi^A_I$ is the parallel propagator on field space evaluated on the inflationary trajectory connecting the field-space coordinates at times $t$ and $t_\epsilon$. The first term in brackets is lowest-order in the slow-roll expansion. The second and third terms exhaust the next-order corrections. For further details of Eq. (B.8) we refer to the literature [202].

**Subhorizon limit.**—We wish to use (B.8) to supply initial conditions for $n^{AB}$ when $k$ is on modestly subhorizon scales for which $|k\eta| \approx k/(aH)$ is in the range 5 to 10. To do so, Ref. [79] retained only the term at lowest order in slow-roll and highest order in $|k\eta|$. Here we retain subleading terms in both expansions.

First, Eq. (B.8) can be used to derive correlation functions involving the momenta $\pi^A \equiv dX^A/dN$. To do so we express $d/dN$ in terms of $\eta$, up to next-order in the slow-roll expansion, using $d/dN = (aH)^{-1}d/d\eta$. Beginning with the unequal-time correlation function, the $N$ derivative on each momentum operator $\pi^A$ can be brought outside the expectation brackets, and the result evaluated by differentiation of the wavefunction factors (B.8). The equal-time limit should be taken only after all such differentiations have been carried out.

Second, after retaining terms in the $ww$ product only up to next-order in the slow-roll expansion, we differentiate with respect to $\ln k$. This determines the element of the spectral matrix corresponding to each correlation function. Note that the evaluation time $\eta$ is fixed and does not vary between $k$-modes. The same is true for the arbitrary evaluation time $t_\epsilon$: neither quantity generates any contribution when differentiated with respect to $k$. Finally, we expand in inverse powers of $k/(aH)$ and retain the leading- and next-order terms. This

---

2In Ref. [202] this expression was given for a flat field manifold. The generalization to a curved field-space was considered in Ref. [66], in which the (B.8) was used in the limit $|k\eta| \to 0$ but this expression did not appear explicitly.
yields
\[
\frac{d}{d \ln k} \langle \delta X^A \delta X^B \rangle' = -\Pi^A_I \Pi^B_J \frac{1}{2ka^2} \left( G^{IJ} + \frac{3 a^2 H^2}{k^2} \left( (2 - \epsilon)G^{IJ} - \frac{M^{IJ}_s}{H^2_s} \right) + \cdots \right),
\]
\[
\frac{d}{d \ln k} \langle \delta X^A \delta \pi^B \rangle' = \Pi^A_I \Pi^B_J \frac{1}{2k} \left( (1 - \epsilon)G^{IJ} - \frac{15 a^4 H^4}{4 k^4} \left( 3\epsilon G^{IJ} - \frac{M^{IJ}_s}{H^2_s} \right) + \cdots \right),
\]
\[
\frac{d}{d \ln k} \langle \delta \pi^A \delta \pi^B \rangle' = \Pi^A_I \Pi^B_J \frac{k}{2H^2 a^4} \left( G^{IJ} + \frac{1 a^2 H^2}{2 k^2} \left( 3\epsilon G^{IJ} - \frac{M^{IJ}_s}{H^2_s} \right) + \cdots \right),
\]
where the notation $\langle \cdots \rangle'$ means that the momentum-conservation $\delta$-function $(2\pi)^3 \delta(k_1 + k_2)$ has been dropped. To write Eqs. (B.9a)–(B.9c) we have combined terms to remove explicit dependence on $t_*$. Therefore $a$, $H$ and $\epsilon$ should be determined at the evaluation time for the correlation function, and at lowest order we can take the Hubble-normalized mass matrix $M^{IJ}_s/H^2_s$ to be evaluated at the same time. Since $\Pi^A_I \Pi^B_J G^{IJ} = G^{AB}$ this means we can take all terms in Eqs. (B.9a)–(B.9c) to be evaluated at the same time and drop the parallel propagator factors.

### B.2 Tensor spectral index

A similar analysis applies for the tensor spectral index. Each polarization behaves like an exactly massless scalar field, up to normalization. Therefore (B.9a)–(B.9c) apply, with $M^{IJ} = 0$. We need only one scalar field, and to account for the normalization of the tensor polarizations we should take $G^{IJ} \rightarrow M^2_P/2$, $G^{IJ} \rightarrow 2/M^2_P$.

We write the momentum for the tensor polarizations as $\pi_s = d\gamma_s/dN$ [75]. Then it follows that
\[
\frac{d}{d \ln k} \langle \gamma_s \gamma_s' \rangle' = -\frac{1}{M^2_P} \frac{1}{k a^2} \delta_{ss'} \left( 1 + \frac{3 a^2 H^2}{2 k^2} (2 - \epsilon) + \cdots \right),
\]
\[
\frac{d}{d \ln k} \langle \gamma_s \pi_s' \rangle' = \frac{1}{M^2_P} \frac{1}{k a^2} \delta_{ss'} \left( 1 - \epsilon - \frac{45 \epsilon a^4 H^4}{4 k^4} + \cdots \right),
\]
\[
\frac{d}{d \ln k} \langle \pi_s \pi_s' \rangle' = \frac{1}{M^2_P} \frac{k}{H^2 a^4} \delta_{ss'} \left( 1 + \frac{3 \epsilon a^2 H^2}{2 k^2} + \cdots \right),
\]
from which the necessary initial conditions can be extracted.
Appendix C

Computational requirements

It was explained in §2 that numerical predictions in the D3-brane model are challenging due to the complexity of the potential. In this Appendix we report in more detail on some of these challenges, in the hope that such details may be useful to researchers interested in incorporating observables derived from three-point functions in their analyses.

Translation step.—For both CppTransport and PyTransport the analysis of a model is a two-step process. First, a model description must be translated into customized C++ code that implements the transport equations for each correlation function. This requires a number of computations involving computer algebra. Second, this code must be processed by the C++ compiler before practical computations can be performed. There are resource implications in each step. Where timings are quoted, they relate to our test machine—an Apple 16" MacBook Pro with 16 GB and an i9-9880H running at 2.30 GHz.

An initial difficulty is that the CppTransport translator does not work successfully out-of-the-box for this model, because the GiNaC library on which it relies performs poorly when computing the inverse conifold metric. This is somewhat surprising given that the inversion is not particularly difficult, but likely reflects the fact that GiNaC is not optimized for matrix operations. Instead, we perform the inversion with SymPy and allow CppTransport to read in the result of the calculation. It is likely that future versions of CppTransport will allow the inverse field-space metric to be specified explicitly as part of the model description. While this reduces the degree of automation in the analysis, it allows more potent tools such as SymPy or Mathematica to be used for those parts of the calculation to which they are best suited.

We find there are benefits to following the same procedure for the components of the connexion and the Riemann tensor. Using SymPy to pre-compute these leads to a significant reduction in the time required for the CppTransport translator to perform common sub-expression elimination (‘CSE’), which is needed to keep time and memory requirements in check during the compile stage. The happens because the CSE algorithm used in CppTransport does not perform well with very large expressions due to design constraints imposed by GiNaC.
With these choices, we find it takes of order $\sim 1$ hr for CppTransport to perform translation on our test machine. While this is not insignificant, the procedure need only be done once.

**Compile step.**—CppTransport produces a core .h file of size $\sim 26$ Mb and an MPI implementation .h file of size $\sim 800$ kb. (For details, see Ref. [106].) In future there may be some scope to reduce the size of the core file by more intelligent pooling of lengthy expressions. Code of this size represents a considerable workload for the compiler, but is still practical with modern systems. We find that Clang is able to compile the model in between 10 min and 20 min, with the exact time depending on the platform and the version of Clang in use. Memory usage peaks between 3 Gb and 4 Gb. We have tested Clang 9 and Clang 10 on Linux and macOS platforms (the latter in the guise of Apple Clang 11 and 12).

GCC is also able to process the model, but it returns significantly slower compile times and has much greater resource utilization. With GCC 10, peak memory usage was recorded in the vicinity of 100 Gb, which may be prohibitive in a virtualized environment or on systems with slow virtual memory. On our test machine, compilation did not terminate in a reasonable time, likely due to performance issues associated with repeated paging to swap. However, we have successfully compiled using GCC on larger machines, including the HPC cluster used to compute our primary catalogue. On these machines the compile time was of order 1 hour. The measurements reported here are not definitive, and are intended only as a guide to relative performance. In particular, compilation performance depends strongly on CPU properties.

Its smaller footprint makes Clang much easier to deploy in a resource-constrained environment such as a virtual machine or Docker container. Subject to this proviso, we find Docker to be a convenient tool to deploy CosmoSIS and CppTransport together.

**Compute time.**—Finally, we consider typical compute times for samples of the inflationary two- and three-point functions. Although there is some variation, computation of a single $k$-mode of the two-point function takes a few seconds. The integration time for modes of the three-point function depends on the momentum configuration. Equilateral and folded configurations take a few hundred seconds, with folded configuration typically being between 1.5 and 2.5 times more expensive than the equilateral. Squeezed configurations are most expensive. For $\beta = 0.9$, corresponding to squeezing $k_3/k_t = 0.05$, the integration is between 10 and 15 times more expensive compared to equilateral configurations. For $\beta = 0.95$, corresponding to squeezing $k_3/k_t = 0.025$, the integration is between 10 and 20 times more expensive compared to equilateral configurations.

In Fig. C.1 we plot the distribution of compute times for the entire set of 55,000 trajectories in our primary catalogue that satisfy the cut $A_s < 10^{-4}$. The individual calculations that contribute to this total are listed in §2.2: note that there is no computation of any squeezed configurations. A majority of trajectories ($\sim 70\%$) complete in less than 1,000 s, and most ($\sim 95\%$) complete in less than 1,500 s. The bimodality is due to trajectories for which we cannot compute the Planck2015 likelihood because there are too few total e-folds to
Figure C.1: Distribution of integration times over our primary catalogue, measured in seconds. We have excluded a tail of rare trajectories up to $10^6$ seconds. The bimodality is due to trajectories for which we cannot compute the Planck2015 likelihood.

sample the power spectrum on the largest necessary scale (see §2.2). For these trajectories we do not need to obtain a separate sample of $P_\zeta$ for use by CLASS, making the calculation substantially shorter.

Although not displayed here, compute times in PyTransport are comparable to (if generally slightly slower than) those reported for CppTransport. Therefore, in this model, inclusion of three-point function observables is likely too slow for use in practical parameter estimation using Markov Chain Monte Carlo methods. It might be possible to compute the primordial two-point function directly from the model, if desired, and providing not too many sample points are used.
Appendix D

Spectral corrections

Figure D.1: Temperature spectral improvements $\Delta C_{TT}^\ell$, as measured from the DX12 143/217 GHz data and simulations. The power spectra are computed directly from the WF maps, and have been binned in multipole intervals of $\Delta \ell = 50$ and weighted by a $L(L+1)$ kernel. In each panel the fractional difference of the fiducial and corrected maps is given in red and green respectively. For reference, two grey shaded bands indicate the 1% and 2% difference contours. Left: Auto-power spectra from 143 GHz. Centre: Cross-power spectra from 143 GHz and 217 GHz. Right: Auto-power spectra from 217 GHz.

Accurate computation of the $N_{L_RD}^{(0)}$ spectral correction described in Section 3.1.3 rely on the Gaussian power between the data and simulations being perturbatively close. In order to ensure this, we add a small amount of Gaussian power to the simulations and/or data. We quantify this difference by measuring the power spectra of the WF data and the ensemble averaged power spectra of WF simulation maps; $\Delta C_{\ell} = C_{\ell, \text{expt}} - \langle C_{\ell}\rangle_{\text{sims}} \approx 0$. The WF maps (rather than the raw maps) provide a better benchmark for assessing the power differences since it is the filtered maps that enter the quadratic estimator. Moreover, in the case of e.g. the (143, 217) GHz analysis where we perform a template projection of the 857 GHz intensity map during the filtering stage, there is a reduction in power that would not otherwise be measured on the raw (unfiltered) maps.

Since we consider the joint modulation reconstruction between map pairs (e.g. see discussion towards the bottom of Section 3.1.2) it is important that we not only correct the Gaussian auto power spectrum within a given map set, but also the cross power between different map sets entering the quadratic estimator. By map sets, we are collectively referring to a given data and its corresponding set of simulations. A concrete example is ensuring that $\Delta C_{\ell} \approx 0$ for 143x143 GHz, 217x217 GHz and 143x217 GHz simultaneously.
We achieve this by solving the system of equations

\[
C(\ell, \Delta \ell) = 0, \quad \text{where } \Delta \ell = \ell_m - \ell_m
\]

where \( P_{\ell m} \in \{T_{\ell m}^{WF}, E_{\ell m}^{WF}, B_{\ell m}^{WF} \} \) and 'a' and 'b' refer to distinct map sets. We want to construct solutions such that \( \Delta C_{\ell}^{ab} = \Delta C_{\ell}^{aa} = \Delta C_{\ell}^{bb} = 0 \). To this end, we assume the following form for the corrections in harmonic space

\[
\begin{align*}
\text{Map set a:} & \quad P_{\ell m}^{a, \text{dat}} & \rightarrow & \hat{P}_{\ell m}^{a, \text{dat}} = P_{\ell m}^{a, \text{dat}} + \theta_{\ell m} + \phi_{\ell m} \\
 & \quad P_{\ell m}^{a, \sim} & \rightarrow & \hat{P}_{\ell m}^{a, \sim} = P_{\ell m}^{a, \sim} + \delta_{\ell m} + \epsilon_{\ell m} \\
\text{Map set b:} & \quad P_{\ell m}^{b, \text{dat}} & \rightarrow & \hat{P}_{\ell m}^{b, \text{dat}} = P_{\ell m}^{b, \text{dat}} + \theta_{\ell m} + \phi_{\ell m} \\
 & \quad P_{\ell m}^{b, \sim} & \rightarrow & \hat{P}_{\ell m}^{b, \sim} = P_{\ell m}^{b, \sim} + \delta_{\ell m} + \epsilon_{\ell m}
\end{align*}
\]

where all of the additive corrections are only correlated with themselves. This means that (by construction) \( \phi_{\ell} \) and \( \epsilon_{\ell} \) fix the cross map set power spectra difference, and the remaining the auto map set power difference.

Substituting the corrected harmonics into our ansatz Eq. (D.1), the power difference can be written in terms of the isotropic power spectrum of the harmonic corrections

\[
\begin{align*}
\Delta C_{\ell}^{ab} &= (C_{\ell}^{a} - C_{\ell}^{b}), \\
\Delta C_{\ell}^{aa} &= (C_{\ell}^{a} - C_{\ell}^{b}) + (C_{\ell}^{a} - C_{\ell}^{a}), \\
\Delta C_{\ell}^{bb} &= (C_{\ell}^{b} - C_{\ell}^{b}) + (C_{\ell}^{b} - C_{\ell}^{b}).
\end{align*}
\]

We can obtain solutions in terms of the combinations given in parentheses (...) by treating them as single variables. The solutions can then be unpacked further by requiring power spectra to be non-negative via a piecewise definition, e.g.

\[
C_{\ell}^{\phi} = \begin{cases} 
|C_{\ell}^{\phi}| & \text{if } (C_{\ell}^{\phi} - C_{\ell}^{\phi}) < 0 \\
0 & \text{if } (C_{\ell}^{\phi} - C_{\ell}^{\phi}) \geq 0
\end{cases}
\]

\[
C_{\ell}^{\epsilon} = \begin{cases} 
(C_{\ell}^{\epsilon} - C_{\ell}^{\epsilon}) & \text{if } (C_{\ell}^{\epsilon} - C_{\ell}^{\epsilon}) \geq 0 \\
0 & \text{if } (C_{\ell}^{\epsilon} - C_{\ell}^{\epsilon}) < 0
\end{cases}
\]

The remaining components \( C_{\ell}^{\delta,a}, C_{\ell}^{\delta,a}, C_{\ell}^{\delta,b} \) and \( C_{\ell}^{\delta,b} \) are obtained in the same way. Since the solutions for the corrective power spectra depend on the empirical power difference, they are in general not smooth functions of \( \ell \). We therefore bin the solutions and interpolate over \( \ell \) in order to construct smooth continuous functions. \(^1\)

Recall that these spectral corrections are built from the power spectra of WF maps. We

\(^1\)A bin width of \( \Delta \ell = 51 \) is used throughout. Note that to minimize boundary effects incurred during binning and interpolation, we use the multipole range 30 \( \leq \ell \leq 2048 \), then retain multipoles 100 \( \leq \ell \leq 2018 \) for final analyses.
want the equivalent result for unfiltered maps on the full sky, such that we can simulate and apply corrections to the raw data and simulations before reprocessing. The required transformation is (approximately) given by

\[
C^{\text{Raw}}_\ell \simeq C^{\text{WF}}_\ell \left(1 + \frac{N_\ell}{C^{\text{Fid.}}_\ell} \right)^2 \frac{B^2_\ell}{f_{\text{sky}}} ,
\]  

(D.11)

where \(B_\ell\) is the beam transfer function and \(N_\ell = \Delta^2 B^{-2}_\ell\) is the approximate isotropic noise power spectrum; for a fixed noise level \(\Delta\). For uncorrelated corrections, we can simulate the corresponding isotropic map corrections directly from Eq. (D.11) using \texttt{healpix}. For correlated components, we should instead simulate harmonic coefficients from the WF spectra, then transform these via an appropriate definition of the instrumental beam and noise (if they differ) for each map set;

\[
a^{\text{Raw}}_{\ell m} \simeq a^{\text{WF}}_{\ell m} \left(1 + \frac{N_\ell}{C^{\text{Fid.}}_\ell} \right) \frac{B_\ell}{\sqrt{f_{\text{sky}}}} ,
\]

(D.12)

before finally transforming the harmonics into real space maps.
Appendix E

Estimated $\tau_{NL}$ from D3-brane inflation

![Graph](image)

**Figure E.1:** Ratio of $\delta N$ non-linearity parameters in D3-brane inflation: $A_{NL} \equiv \tau_{NL}/(6f_{NL}/5)^2$. *Upper:* ‘Sample A’, satisfying Planck constraints on $n_s$. *Lower:* ‘Sample B’ with $f_{NL}^{\text{eq}} \sim -5$. 


[38] Julien Lesgourgues. The cosmic linear anisotropy solving system (CLASS) i: Overview, 2011.


[204] Edmund J. Copeland, Edward W. Kolb, Andrew R. Liddle, and James E. Lidsey.


[257] Neha Anil Kumar, Gabriela Sato-Polito, Marc Kamionkowski, and Selim C. Hotinli. Primordial trispectrum from kSZ tomography. 5 2022.


