The Effective Field Theory Approach to Fluid Dynamics, Modified Gravity Theories, and Cosmology

Junpu Wang

The effective field theory approach is powerful in understanding the low energy phenomena without invoking the UV degrees of freedom. We construct a low energy Lagrangian for ordinary fluid systems (in constrast to superfluid), pure from symmetry considerations and EFT principles. The dynamical fields are the Goldstone excitations, associated with spontaneously broken spacetime translations. It is organized as derivatively coupled theory involving multiple scalar fields. This formalism enables us to study fluid’s quantum mechanical properties and dissipative effects. Cosmological models can be built by naturally coupling the fluid EFT to gravity.

From the EFT point of view, GR is the unique low energy theory for the spin-2 graviton field and any infrared modification corresponds to adding new degrees of freedom. We focus on two popular classes of modified gravity models, — the chameleon like theories and the Galileon theory, — and perform a few reliability checks for their qualifications as modified gravity theories.

Furthermore, guiled by the EFT spirit, we develop a cosmological model where primordial inflation is driven by a ‘solid’, defined, in a similar manner as the EFT of fluid. The symmetry breaking pattern differs drastically from that of standard inflationary models: time translations are unbroken. This prevents our model from fitting into the standard EFT description of adiabatic perturbations, with crucial consequences for the dynamics of cosmological perturbations, and exhibits various unusual features.
# Table of Contents

1 Introduction  

2 EFT Construction for Fluids  
   2.1 Effective Field Theory for Perfect Fluids  
      2.1.1 Fluids with No Conserved Charge  
      2.1.2 Fluids with Conserved Charge  
      2.1.3 Hydrodynamics and Thermal Relations  
      2.1.4 The Goldstone Field Action  
      2.1.5 Duality Between Irrotational Ordinary Fluid and Superfluid  
   2.2 EFT for Imperfect Fluids  
      2.2.1 The general idea  
      2.2.2 Low frequency, long distance behavior of correlators  
      2.2.3 Dissipative couplings: $S_{int}$ to linear order  
      2.2.4 Rediscovering Kubo relations  
   2.3 A Glance at the Quantum Mechanics of Perfect Fluids  
      2.3.1 A Naive Deformation  
      2.3.2 Scattering Amplitude of Some Simple Processes — the Vortex Strong Coupling  
      2.3.3 The infrared situation  
   2.4 Application of the Effective Fluid Model to Cosmology  
      2.4.1 Perfect Fluids in a Cosmological Background  
      2.4.2 The Conserved Curvature Perturbation to Non-linear Order
5 Conclusions 115

Bibliography 119

A Linear couplings of a $U(1)$ Goldstone 125

B Phase Space for Phonons 129

C Expressions for $Z^\mu_{\nu}$ in the Galileon Theory 131

D Further Details of the Galileon Stability Analysis 132

E Unitary Gauge vs. Spatially Flat Slicing Gauge for Solid Inflation 135

F Full trilinear action in SFSG for Solid Inflation 137

G Time Dependence of Background Quantities in Solid Inflation 139
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Feynman diagram representation of Eqn. (2.39)</td>
<td>20</td>
</tr>
<tr>
<td>2.2</td>
<td>Tree Diagrams for LL→ LL Scattering</td>
<td>36</td>
</tr>
<tr>
<td>2.3</td>
<td>Tree Diagrams for TT→ TT Scattering</td>
<td>37</td>
</tr>
<tr>
<td>4.1</td>
<td>The shape of non-gaussianities in solid inflation model</td>
<td>105</td>
</tr>
</tbody>
</table>
Acknowledgments

First and foremost, I want to thank my parents.

Although having their only child tens of thousand miles away, they never complained. On the contrary, they constantly encourage and support me to pursue what I like. They also have influenced, to some extent, the way I am doing research today. Without them, I can hardly imagine that I could achieve a Ph.D degree some day.

I can not overestimate my gratitude towards my thesis supervisor Professor Alberto Nicolis. He is my mentor; he is the person who leads into the palace of modern physics; he forges many of my viewpoints about quantum field theory and about gravity. But more often, he is like a close friend: I can walk in his office asking random questions and even making jokes. I would not expect any more from a professor than I get from Alberto.

I want to always express my deepest gratitude towards my co-supervisor Professor Lam Hui. It is my honor to get guidance from two young, brilliant scientists during my doctoral study. Lam has an incredible ability to master knowledge from a vast of amount of fields. He is an example, stimulating me to dabble in other areas that I am not familiar with.

I owe a debt of gratitude to Professor Justin Khoury, who has been a tremendous help to me in our collaboration and in the course of my applying for a postdoc job. I appreciate the encouragement and the advice about future plan that he gave.
Friends have always been a big part throughout my life; especially in the journey of
pursuing my Ph.d degree, they are my listeners and cheerleaders. Shan Gao, my college
roommate, colleague and best friend, thank you for all the conversations we have, about
physics, about future and about all kinds of random things. Yaocheng Fei, Xiaoqiang
Zhong and Shuyu Zhou, thank you buddies! Despite the long distance and infrequent
communication, every time I talked to/saw you, I felt most relaxing and pleasantest.

I remember 6 years ago when I came to Columbia, it was my first time to leave my
home country and my first time to live alone it a foreign city. Fortunately, I met lots of
new friends here. To Kehui Coffee Xiang, Yinsheng Guo, Dazhi Tan, Wenbo Li, Zhou Ye
and our captain Tianxia Jia, you are making my life in this city delightful. Thank you for
the basketball games we played, the basketball/soccer/American football/baseball games
we watched and the dinners we had together. Nothing works better than these gathering
events to dispel loneliness.

I am grateful to my colleagues (also friends). I want to thank Solomon Endlich — the
person who I have done most of my projects with and also the person who offered me
enormous help in overcoming the culture gap. I would like to thank Andrej Ficnar, Xiao
Xiao, Bart Horn, Riccardo Penco, Rachel Rosen for all the discussions we had, from which
I benefited tremendously. To Professor Eduardo Ponton, thanks for letting me sit in your
QFT and particle physics classes and for patiently answering my confusions in or outside
the scope of classes. To Zhongjie Lin and Hantao Yin, thanks for all your assistance on
programming, algorithms and computer hard/soft-wares. There are many more people I
want to thank; at least allow me to list their names here: Chris Chan, Qi Liu, Yujiao Chen,
Jianglei Yu, Feng Tian, Daiqian Zhang, Denial Hu (my two year roommate), Alex Chen,
Junichi Okamoto, Hung The Dang, Dashun Wang, Lan Gong . . .

I am also grateful to my advisors in college Professor Zhifang Lin and Professor Rukeng
Su. The former suggested me learn Mathematica, which now becomes my main toolkit to
attack hard analytic problems and/or non-analytic problems. The latter kept reminding
me the hardness in the theoretical physics journey, which makes me never too optimistic
about what I have achieved.

I would like to show my gratitude to the coordinators in the department of physics: Lalla
Grimes (retired in 2009), Rasma Mednis (May she be at peace in heaven) and Randy Torres, who have kindly helped me deal with all sorts of paperwork, especially the documents for visa application every time when I travel. I would also thank our finance officer Yasmin Yabyabin and Lydia Argote for their taking care of my tedious reimbursement issues.

I would like to mention the name of my favorite basketball player, Steve Nash — an not so physically gifted player in NBA who has won the Most Valuable Player award twice — whose experience reminds me how important hard-working and persistence could be in achieving success.

Last but not least, to my beloved Tianyan Hu, thank you for letting me pursue my dream. Your understanding and support are the best gift I have ever had.
TO MY PARENTS XIAOFENG MA AND YULIN WANG
Chapter 1

Introduction

The dream of physics is to establish a theory that could possibly explain everything in nature. However, even if such a *theory for everything* is found at some point, in practice one will not expect to be able to use it for phenomena at all physical scales. For instance, it is believed that at classical level Einstein’s General Relativity (GR) is the correct gravity theory. But when one wants to know how a ball flies after it is thrown out of a window, we only need to invoke the Newtonian gravity theory and the Newtonian mechanics, without even mentioning concepts such as metric, curvature of the spacetime, etc. . . — which are at the center of GR. Said differently, in solving mechanical problems on earth, the Newtonian gravity theory — as an *effective theory* of GR — works more efficiently.

Another (maybe more up-to-date) example is from Quantum Field Theory (QFT), which successfully provides the framework describing the various interactions between subatomic particles. In computing the 4 lepton scattering amplitude at low external energy, it is more convenient to use Fermi’s effective four-fermion theory rather than the full electroweak theory (i.e. the $SU(2) \times U(1)$ gauge theory). The reason for this is that at low external momentum, exchanging virtual $W$ or $Z$ bosons is just a local interaction.

In fact, in the context of QFT, *effective field theory* does more than simplifying the computation; it is closely related to some fundamental structure of QFT (e.g. the Renormalization Group (RG) flow) and changes radically the viewpoint on certain aspects that were considered to be essential (e.g. Renormalizability).

It was first noticed by Wilson that when the external energy $E$ of some quantum process
are lower that some mass scale $\Lambda$, the Ultra-Violet (UV) degrees of freedom above $\Lambda$ will not be excited. So one can ‘integrate out’ these UV d.o.f.’s, and obtain a low energy effective theory that captures completely the dynamics of the low energy sector. The imprint from the UV excitations in this low energy effective theory is a bunch of (usually infinite) high dimension interaction terms (thus non-renormalizable), each suppressed by powers of $E/\Lambda$.

As one can see immediately, this Wilsonian integrating out procedure breaks down as $E$ is comparable to $\Lambda$, since then the heavier excitations are awakened.

Long after the birth of QFT when physicists were struggling to understand the intrinsic divergences in QFT, it was believed that only renormalizable theories made sense. However, from the EFT point of view, the renormalizability seems irrelevant: the fact that a non-renormalizable theory breaks down at certain energy level $\sim E_{\ast}$ simply implies that this theory is incapable of describing the physics at distance smaller than $\sim E_{\ast}^{-1}$. This should not be surprising, since probably we haven’t included into the theory the heavier d.o.f.’s that will become relevant at short distance. In that sense, the non-renormalizability implies an incompleteness in description, rather than a pathology of the theory.

On the other hand (and maybe more strikingly), the non-renormalizable theories, viewed as EFT’s, are able to make predictions on low energy phenomena. The reason is the following: as we mentioned, all the high dimension operators are suppressed by powers of $E/\Lambda$. Therefore, at low energy $E$ and to any required precision $\epsilon$, there are a only finite number of coefficients we need to determine — i.e. for $n \in \mathbb{Z}^+$ such that $(E/\Lambda)^n \gtrsim \epsilon$, either by knowing the underlying UV-complete theory or by measurements from experiments.

After that, we can use this theory to make predictions at the energy level $\sim E$, the same way as we use any renormalizable theories.

The Wilsonian procedure — i.e. the way of obtaining a low energy EFT via integrating out the UV d.o.f. — is referred to as top-down; we need know (or at least have some idea) what the UV theory is. However, in many practical circumstances, we don’t\footnote{Or in some other cases, even though we know the UV theory, the ‘integrating out’ procedure will be much more complicated than the normal perturbative RG flow. The fluid system and/or hydrodynamics are just such a example.} But if we are not so ambitious and content ourselves with the physics in the low energy regime, we can
adopt the bottom-up method — the so called \textit{EFT approach} or \textit{EFT construction}, — which will be the main subject of this thesis. Indeed, since not every piece of the full, possibly complicated, UV theory will play a vital role in low energy physics, then why shall we bother to obtain it? In that sense, the EFT approach, although less aesthetically elegant, seems more economic.

Now what are the rules of the game? Suppose we want to describe the low energy/long wavelength phenomena of some physical system. First we need to determine the field contents that could be used as the low energy d.o.f.'s. Then we specify the symmetry properties of the system, and write down all possible terms in Lagrangian that are constructed from the IR fields and are compatible with symmetries. Since usually the UV completion of such an EFT remains unknown, the coefficients in the EFT has to be determined by comparing predictions in this theory with known results from experiments.

This EFT approach turns out to be useful in many areas and can be apply to a variety of physical systems. The first application we will investigate is to recast fluids into a low energy effective field theoretical description, with an emphasis on its internal and spacetime symmetries, their spontaneous breaking pattern, the associated Goldstone excitations and their interactions, the derivative expansion, and in general on the systematics of the EFT program \cite{1-6}. We will review this EFT construction briefly in Section \ref{sec:2.1} where, we will content ourselves the leading order in the derivative expansion, which corresponds to focusing only the non-dissipative properties.

Recall from the classical fluid dynamics that dissipation appears in the gradient expansion of hydrodynamics as a first order correction to the perfect fluid equations, which are the continuity and the (relativistic generalization of the) Euler equations. So, naturally we expect that when we continue our EFT construction to first order in derivatives, we reveal dissipation. Technically, including dissipative effects in the EFT formulation is non-trivial, since the Hamiltonian of the system is always a conserved quantity if there is not explicit time dependence. It is our task in Section \ref{sec:2.2} to illustrate how to do this systematically.

This approach offers a number of advantages over the traditional one — which starts from the equations of motions. For instance, using the zeroth order Lagrangian for fluids (i.e. perfect fluids), we can now compute straightforwardly cross sections associated with the
processes of sound waves and vortices scattering off each other, via the Feynman diagram techniques borrowed from QFT. Remarkably, via these computations, we found that quantum mechanically perfect fluids exhibit peculiar features associated with the vortex degrees of freedom. This stems from the vortex modes’ lack of gradient energy, which makes them not behave like normal quantum fields. This is the topic of Section 2.3.

Another advantage of having a Lagrangian description for fluids is to make it easy to couple the fluid system to other sectors, e.g. gravity. We just replace the flat Minkowskian metric \( \eta_{\mu\nu} \) with the dynamical one \( g_{\mu\nu} \). Such a construction offers explicitly a cosmological action, with the matter contents described by the EFT of fluids and makes it really easy to tackle certain questions. For instance, in Section 2.4, we study how to construct the nonlinear curvature perturbation \( \zeta \) — the linear definition was first proposed by Bardeen \cite{7} — which is expected not to evolve outside the horizon.\footnote{c.f. for instance \cite{8} for a proof of the super-horizon conservation of the linear curvature perturbation.} Many studies have been done to extend the construction of the conserved \( \zeta \) to non-linear level in different contexts \cite{9–19}. But we will provide a novel and somewhat neater way, thanks to the EFT of fluids.

In addition to mundane systems like fluids (or solids), this EFT approach can also be applied to study gravity theories. Although GR is widely accepted as the correct gravity theory at classical level, it can not be the \textit{ultimate} theory for gravity, because of its derivatively coupled (hence non-renormalizable) nature. As we argued previously, this is hardly a problem for low energy physics, but rather implies that we should really regard GR as a low energy quantum EFT. More remarkably, as we will argue, it is the \textit{unique} effective field theory for a massless spin-2 field at low energy.

Despite its robustness from the viewpoint of theoretical considerations and its agreement with local gravity tests, there are attempts to modify GR at the infrared scales. The main motivation comes from the discovery of the universe’s accelerating expansion. The \( \Lambda \)-Cold Dark Matter (\( \Lambda \)CDM) standard model of cosmology — which is based on Einstein’s gravity theory — attributes this to the presence of some unknown substance in the universe, the so called \textit{dark energy}. Since the energy density of this species will not dilute as the universe expands, people also refer it as the \textit{Cosmological Constant} (CC). There is a \textit{naturalness} problem associated with the CC: roughly speaking, the value for CC needed
to drive the accelerating expansion is fascinatingly small compared to that estimated from theory (c.f. [21,22] for a review of the CC problem). Of course it will be theoretically more interesting that the (small) CC is actually due to some dynamical mechanism that modifies GR at cosmological scales, rather than to a new matter component in the universe.

The simplest possibility is to add to GR an extra scalar field. There are a vast number of proposals along this direction, among which we will focus on two — the chameleon theory [23,24] and the Galileon theory [26]. The are both theoretically well motivated and observationally viable. By the latter we mean that these theories intrinsically have screening mechanisms (which will be discussed in detail in Section 3.2) to ‘hide’ the scalar force so that the consistency with local gravity experiments is ensured. Further checks about the reliability and about the usefulness of these theories [27, 28] will be performed in Section 3.3. 3.4.

It is tempting to apply this EFT approach to cosmology. In fact there are studies on constructing inflationary models using this EFT approach [29,30]. The basic idea is the following: in the usual inflationary scenarios, the matter fields $\psi_m$ feature time-dependent cosmological background solutions $\bar{\psi}_m(t)$, which spontaneously break time translations. As a result, there is one fluctuation mode $\pi(x)$ that can be identified with the associated Goldstone excitation. Roughly speaking, it can be thought of as an in-sync perturbation of all the matter fields, of the form

$$\psi(x) = \bar{\psi}_m(t + \pi(x)) \simeq \bar{\psi}_m(t) + \partial_t \bar{\psi}_m(t) \cdot \pi(x).$$ (1.1)

When coupling to gravity is taken into account, such a mode describes adiabatic perturbations. As usual for Goldstone bosons, the spontaneously broken symmetry puts completely general, non-trivial constraints on these perturbations’ dynamics. This property is at the basis of the model-independent approach that goes under the name of “effective field theory of inflation”, whose tenets are particularly compelling since they encompass—at first glance—all cosmological models: cosmology is about time-dependent, homogeneous, and isotropic field configurations.

However, as we will show in Chapter 4 there are other possibilities [25]. In our case, we will be dealing with matter fields featuring time-independent, $\vec{x}$-dependent background solutions. Apparently, this contradicts two facts about inflationary cosmology:
1. The universe is homogeneous and isotropic;

2. In an expanding universe physical quantities depend on time and, more to the point, that one needs a physical ‘clock’—a time-dependent observable—to tell the universe when to stop inflating.

As for item 1: $\bar{\xi}$-dependent solutions can be compatible with the homogeneity and isotropy we want for cosmological solutions and for the dynamics of their perturbations, provided extra symmetries are imposed. For instance, to get an FRW solution for the gravitational field, we need an homogeneous and isotropic background stress-energy tensor. This can arise from matter fields that are not homogeneous nor isotropic, provided there are internal symmetries acting on the fields that can reabsorb the variations one gets by performing translations and rotations. The simplest example is that of a scalar field with a vacuum expectation value

$$\langle \phi \rangle = \alpha x.$$  \hfill (1.2)

Such a configuration breaks translations along $x$. However, if one postulates an internal shift symmetry $\phi \to \phi + a$, then the configuration above is invariant under a combined spacial translation/internal shift transformation. As we will see, this is enough to make the stress-energy tensor and the action for small perturbations invariant under translations. To recover isotropy as well, one needs more fields, and more symmetries. For instance—in fact, this is the case that we will consider—one can use three scalar fields $\phi^I(x)$ ($I = 1, 2, 3$), with internal shift and rotational symmetries

$$\phi^I \to \phi^I + a^I, \quad a^I = \text{const},$$ \hfill (1.3)

$$\phi^I \to O^{IJ}\phi^J, \quad O^{IJ} \in SO(3),$$ \hfill (1.4)

so that the background configurations

$$\langle \phi^I \rangle = \alpha x^I$$ \hfill (1.5)

are invariant under combined spacial translation/internal shift transformations, and under combined spacial/internal rotations. As we will review in Section 4.1 such a system has
the same dynamics as those of the mechanical deformations of a solid—the phonons. In this sense, the cosmological model that we are putting forward corresponds to having a solid driving inflation. If this interpretation causes the reader discomfort—in particular, if having a solid that can be stretched by a factor of $\sim e^{60}$ without breaking sounds implausible—one should think of our model just as a certain scalar field theory. As we will see, the structure of such a theory is the most general one compatible with the postulated symmetries—and the impressive stretchability we need can also be motivated by an approximate symmetry—so that from an effective field theory standpoint, ours is a perfectly sensible inflationary model. From this viewpoint, the fact that the solids we are used to in everyday life behave quite differently—quantitatively, not qualitatively—seems to be an accident: they lack the ‘stretchability symmetry’.

As for apparent contradiction number 2: In our model the role of the physical clock will be played by the metric. More precisely, it will be played by (gauge invariant) observables, made up of our scalars and of the metric, like for instance the energy density or the pressure. These can depend on time even for purely space-dependent scalar backgrounds, because in the presence of a non-trivial stress-energy tensor, the metric will depend on time, in a standard FRW fashion. Doesn’t this correspond to a spontaneous breakdown of time translations too? At some level it is a matter of definition, but we will argue in Section 4.8 that the operationally useful answer is ‘no’, in the sense that there is no associated Goldstone boson, and that one cannot apply to our case the standard construction of the effective field theory of inflation as given in [30].

Formal considerations aside, our peculiar symmetry-breaking pattern has concrete physical implications, with striking observational consequences. For instance, it predicts a three-point function for adiabatic perturbations with a ‘shape’ that is not encountered in any other model we are aware of. Its overall amplitude is also unusually large, corresponding to $f_{NL} \sim \frac{1}{\epsilon} \frac{1}{c_s^2}$.

It is worth mentioning that we have been using (and will be using) a somewhat mislead-

---

3Perfect fluids can not drive inflation, for the reason that we will explain in Section 4.2. Roughly speaking, it is impossible for the fluid EFT to remain consistent (i.e. weakly coupled) while at the same time to provide an energy momentum tensor for an acceleratingly expanding universe.
Chapter 1. Introduction

ing, but fairly standard, language: when spontaneously broken symmetries are gauged, the associated would-be Goldstone bosons are not in the physical spectrum—rather, they are ‘eaten’ by the longitudinal polarizations of the gauge fields. In fact, there is a gauge, the so-called unitary gauge, in which the Goldstone fields are set to zero. In our case we will be dealing with spontaneously broken translations and rotations, and when gravity is dynamical, these are gauged. In unitary gauge one can set the scalars to their vevs (1.5), and have the corresponding excitations show up in the metric. The Goldstone language is still useful though, in that it captures the correct high energy/short distance dynamics of these excitations. For massive gauge theories, this statement goes under the name of “the equivalence theorem”. For cosmological models, it is just the statement that at sub-cosmological distances and time-scales, in first approximation one can neglect the mixing between matter perturbations and gravitational ones. We hope the Goldstone boson nomenclature will be more useful than misleading.

During the doctoral studies, the author has also been working on other projects, which will not be included in this thesis. In [31], we proved that there was no stable stationary soliton solution in Galileon theories. The same argument can be applied to a different class of derivatively coupled theories — e.g. EFT for superfluids, fluids, solids and some k-essence models —, and therefore these theories do not possess stable soliton solutions either. In [32], we performed some explicit computations to confirm the dual theorem about perfect fluids, proposed in [1]: a perfect ordinary fluid free of vortex modes is equivalent to a superfluid system. In particular, we computed the amplitude for 2 to 2 sound wave scattering using the ordinary fluid EFT and the superfluid EFT, respectively, and we showed that the results were equal.
Chapter 2

EFT Construction for Fluids

2.1 Effective Field Theory for Perfect Fluids

2.1.1 Fluids with No Conserved Charge

In this subsection, we first focus on constructing a Poincaré invariant low energy effective theory for an ordinary perfect fluid system with vanishing chemical potential $\mu = 0$. We specify as the long wavelength degrees of freedom the comoving coordinates of fluid volume elements, parametrized by $\phi^I$, with $I = 1, 2, 3$. So at any fixed time $t$, the physical position occupied by each volume element is given by $\vec{x}(\phi^I, t)$. In this description, known as the Euclidean description, the physical spatial coordinates $x^i$ serve as dynamical fields while $t$ and $\phi^I$ are analogous to world sheet coordinates.

However, it is often more convenient to use the inverse functions $\phi^I(\vec{x}, t)$ \footnote{For space filling fluid, there is a diffeomorphism between the comoving coordinates $\phi^I$ and physical spatial coordinates $x^i$ so that for any fixed time $t$, the inverse functions $\phi^I(\vec{x}, t)$ exist.} as dynamical degrees of freedom (known as the Lagrangian description), since the spacetime symmetry can be straightforwardly implemented — we simply demand $\phi^I$ to transform as scalars under Poincaré transformations. Also we are allowed to choose the comoving coordinates in such a manner that when the fluid system is at rest, in equilibrium and in a homogeneous state at some given external pressure, $\phi^I = x^I$ — in the field theoretical language, this is
equivalent to specifying the ground state of our theory to have

\[ \langle \phi^I \rangle = x^I \]  \hspace{1cm} (2.1)

What are the other symmetries, in addition to the Poincaré invariance, required to make the system behave like an ordinary fluid? Notice that our choice of the ground state \( (2.1) \) breaks both spatial translational and rotational invariance. In order for the energy momentum tensor of the system in equilibrium to remain homogeneous and isotropic, as is indeed the case for a fluid, we shall impose internal symmetries to compensate the spontaneously broken spacetime symmetries. More precisely, we demand that the theory be invariant under such internal transformations (i.e. all fields evaluated at the same spacetime point) as:

- \( T_i : \phi^I \rightarrow \phi'^I = \phi^I + a^I, \quad a^I \text{ constant} \) \hspace{1cm} (2.2)
- \( R_i : \phi^I \rightarrow \phi'^I = O^I_J \phi^J, \quad O^I_J \in SO(3) \) \hspace{1cm} (2.3)

It is easy to show that the background configuration \( (2.1) \) is invariant under the diagonal translation and rotation, respectively, which are defined as a linear combination of the spatial and internal transformations : \( T_d \equiv T_s + T_i \) and \( R_d \equiv R_s + R_i \); it is these residual symmetries that correspond to the homogeneity and isotropy of our background configuration.

Moreover an ordinary fluid is insensitive to incompressional deformations — it costs no energy to displace fluid volume elements if they are not compressed or dilated. Expressed in terms of a symmetry requirement, we demand that the theory be invariant under volume preserving differomorphisms of the comoving coordinates, defined as

\[ D : \phi^I \rightarrow \xi^I(\vec{\phi}), \quad \text{with} \quad \det \frac{\partial \xi^I}{\partial \phi^J} = 1. \]  \hspace{1cm} (2.4)

So now we are ready to construct the effective action for an ordinary fluid from the \( \phi^I \)'s, compatible with the symmetry properties mentioned above. It is organized as a derivative expansion. Since the internal translation symmetry \( (2.2) \) mandates each field to be accompanied by at least one derivative, at the leading order in the derivative expansion, the only invariant is

\[ B \equiv \det(B^{IJ}), \quad \text{with} \quad B^{IJ} = \partial_\mu \phi^I \partial^\mu \phi^J \]  \hspace{1cm} (2.5)
and hence the effective action takes the form of

$$S_{\text{Fluid}} = \int d^4x \, F(B)$$  \hspace{1cm} (2.6)

where $F$ is a generic function, which, as we will see, characterizes the equation of state of the fluid in question.

To illustrate that this effective action indeed describes a perfect fluid, we need to check that the (conserved) energy moment tensor takes the famous form:

$$T_{\mu\nu} = (\rho + p) u_\mu u_\nu + p \eta_{\mu\nu}$$  \hspace{1cm} (2.7)

Before doing that let’s work out the four-velocity field $u^\mu(x)$ for the fluid system in our field theoretical language, which by definition is such that

$$0 = \frac{d}{d\tau} \phi^I(x) \equiv u^\mu(x) \partial_\mu \phi^I(x)$$  \hspace{1cm} (2.8)

where $\tau$ parametrizes the streamline and the derivative with respect to $\tau$ vanishes because the comoving coordinate (or the label) of each fluid volume element is fixed. By solving the above equation for $u^\mu$, we obtain

$$u^\mu = -\frac{1}{\sqrt{B}} \epsilon^{\mu\alpha\beta\gamma} \partial_\alpha \phi^1 \partial_\beta \phi^2 \partial_\gamma \phi^3$$  \hspace{1cm} (2.9)

where $\epsilon$ is the 4d Levi-Civita symbol, with the convention $\epsilon_{0123} = -\epsilon^{0123} = 1$. The normalization and overall sign of $u^\mu$ are chosen such that $u^\mu u_\mu = -1$ and $u^0 > 0$.

The energy momentum tensor following from the effective action (2.6) reads

$$T_{\mu\nu} = -2F'(B)B(B^{-1})_{IJ} \partial_\mu \phi^I \partial_\nu \phi^J + \eta_{\mu\nu} F(B).$$  \hspace{1cm} (2.10)

With the aid of (2.18), the energy momentum tensor above indeed can be recast into the perfect fluid form (2.7), if we identify $\rho = -F(B)$ and $p = F(B) - 2F'(B)B$. This also justifies our preceding claim that the generic function $F$ determines the equation of state for fluids. For instance, for an ultra-relativistic fluid with $p = \rho/3$, one has $F(B) \propto B^{2/3}$.

In fact, one can show that the conservation of the energy momentum tensor (2.10) encodes all information about the (classical) dynamics of the system, in the sense that it is equivalent to the equations of motion for the three scalar fields.
\[ \partial_{\mu} T^{\mu\nu} = 0 \iff \partial_{\mu} \left( 2F'(B)B(B^{-1})_{IJ} \partial^{\mu} \phi^{J} \right) = 0, \quad I, J = 1, 2, 3. \] (2.11)

On the other hand, this also confirms our aforementioned claim that the system specified by \( \mathcal{L} = F(B) \) contains no conserved current.

### 2.1.2 Fluids with Conserved Charge

For perfect fluids system with nonzero chemical potential, there will be other conserved quantities, such as the electric charge, the number difference of baryons and antibaryons etc., and hence the dynamics of the system is governed by the equations of conservation of the energy and momentum as well as by that of the particle number: \( 0 = \partial_{\mu} j^{\mu} = \partial_{\mu} (nu^{\mu}) \), where \( n \) is the particle number density.

Thus to model such a system in effective field theoretical language, we need more field contents to represent the particle number symmetry. It turns out he most economical addition is a real scalar \( \psi(x) \), whose background configurations is chosen along the time direction \( [4] \):

\[ \langle \psi(x) \rangle \propto t \] (2.12)

Furthermore, to make the low energy effective field theory for the \( \phi^{I} \)'s and \( \psi \) have the correct fluid-like behavior, — in particular for the EFT to have a conserved current \( j^{\mu} \propto u^{\mu} \) —, we should require that, in addition to the internal symmetries (2.2) (2.3), the theory be invariant under the comoving-position-dependent \( U(1) \) shift:

\[ \psi \rightarrow \psi + \mathcal{F}(\phi^{I}). \] (2.13)

Notice that at the lowest order in derivative expansion, the Lorentz scalars that are invariant under the internal symmetries (2.2) (2.3) and (2.13) can only be constructed from \( B \) (c.f. (2.5)) and \( \sqrt{B} u^{\mu} \partial_{\mu} \psi \) (\( u^{\mu} \) is still defined as in (2.18)). The reason why the latter is invariant under the chemical shift follows from the orthogonality of \( u^{\mu} \) and \( \partial_{\mu} \phi^{I} \). Thus we write the effective action as

\[ S = \int d^{4}xF(b, y), \quad \text{with} \quad b \equiv \sqrt{B}, \quad y \equiv u^{\mu} \partial_{\mu} \psi. \] (2.14)
Once again, \( F \) is a generic function and is related to the equation of state of the fluid system.

We can repeat the procedure of the last subsection and show that the energy momentum tensor following from the above action does take the form of (2.7), if we identify \( \rho \) and \( p \) with

\[
\rho = F_y y - F, \quad p = F - F_b b. \tag{2.15}
\]

where \( F_y \equiv \partial F/\partial y \) and \( F_b \equiv \partial F/\partial b \).

We can also compute the Noether current of the global \( U(1) \) transformation — Eqn. (2.13) with \( F \) set to be a constant,

\[
j^\mu = F_y u^\mu \tag{2.16}
\]

As was mentioned before, we expect this Noether current to represent the conserved particle current. That implies that we need also to identify \( F_y \) with the particle number density of the fluid:

\[
n = F_y \tag{2.17}
\]

### 2.1.3 Hydrodynamics and Thermal Relations

Now let’s make contact with the standard hydrodynamic and thermodynamic description of a fluid with a conserved particle current.

Notice that we have found the fluid’s energy density and pressure in (2.15) and its particle number density in (2.17). Moreover, we have an identically conserved current \( J^\mu = \sqrt{B} u^\mu \),

\[
\partial_\mu J^\mu = -\partial_\mu (\epsilon^{\mu \alpha \beta \gamma} \partial_\alpha \phi^1 \partial_\beta \phi^2 \partial_\gamma \phi^3) = 0, \tag{2.18}
\]

which should be naturally interpreted as the entropy current in the context of our non-dissipative (perfect) fluid. That is \( J^\mu = su^\mu \); and thus \( b \) should be understood as the entropy density \( s \):

\[
s = b. \tag{2.19}
\]

The other thermodynamic quantities such as the temperature \( T \) and the chemical potential \( \mu \) can be found by imposing the thermodynamic identities:

\[
\rho + p = Ts + \mu n, \quad d\rho = T ds + \mu dn. \tag{2.20}
\]
CHAPTER 2. EFT CONSTRUCTION FOR FLUIDS

14

It follows immediately that

\[ T = -F_b, \quad \mu = y. \]  

(2.21)

2.1.4 The Goldstone Field Action

In this subsection, we will consider small fluctuations in fluids. For simplicity, we only focus on fluids with vanishing chemical potential. The perturbations about the homogeneous equilibrium background configuration \([2.1]\) can be parametrized as

\[ \phi^I = x^I + \pi^I(x). \]  

(2.22)

These \( \pi^I \) fields are associated with Goldstone excitations.

Now we want to work out explicitly the effective action for those Goldstone fields. It turns out convenient to rewrite the original Lagrangian \((2.6)\) as \([2]\)

\[ \mathcal{L} = -w_0 f(\sqrt{B}) \]  

(2.23)

where \( w_0 = -2F'(1) = (\rho+p)_{B=1} \) is the ground-state’s enthalpy density, and \( f \) is normalized accordingly, so that \( f'(1) = 1. \) Note that now the derivatives of \( f \) are with respect to \( \sqrt{B}. \)

Expanding the above action to quadratic order, we find that

\[ \mathcal{L}^{(2)} = \frac{1}{2} w_0 \left( \pi_L^2 - c_s^2 (\vec{\partial} \pi_L)^2 \right) + \frac{1}{2} w_0 \dot{\pi}_T^2 \]  

(2.24)

where \( \pi_L \) and \( \vec{\pi}_T \) are the longitudinal (curl-free) and transverse (divergence-free) components of \( \pi^I:\)

\[ \pi^I = \frac{\partial^I}{\sqrt{-\partial^2}} \pi_L + \pi_T^I \]  

(2.25)

The (squared) speed of sound \( c_s^2 \) is given by

\[ c_s^2 = \frac{dp}{d\rho} \bigg|_{B=1} = \frac{2F''(B)B + F'(B)}{F'(B)} \bigg|_{B=1} = f''(1) \]  

(2.26)

Indeed, we see that only the longitudinal Goldstone field \( \pi_L \) admits the standard propagating mode at a finite speed \( c_s \), with a dispersion relation \( \omega = c_s k \), while the other Goldstone field \( \vec{\pi}_T \) has a degenerate dispersion relation \( \omega = 0 \) and thus does not propagate. For this reason, we usually interpret \( \pi_L \) as the sound wave d.o.f. and \( \vec{\pi}_T \) as the vortex d.o.f.: vortices do not “propagate” at large distances.
Notice that after spontaneous breaking, the spacetime symmetries get mixed with the internal ones, so the three Goldstone fields $\pi^I$ transform as a vector field under the diagonal $SO(3)$. And after the decomposition (2.25), $\pi_L$ can be regarded as a scalar field and $\vec{\pi}_T$ as a transverse vector field.

We can further expand the Lagrangian (2.23) to higher orders in Goldstone fields to investigate their interactions. Up to fourth order, we get

$$
L \simeq w_0 \left\{ \frac{1}{2} c_s^2 [\partial \pi] [\partial \pi^2] - \frac{1}{6} (3 c_s^2 + f_3) [\partial \pi]^3 + \frac{1}{2} (1 + c_s^2) [\partial \pi] \dot{\vec{\pi}}^2 - \dot{\vec{\pi}} \cdot \partial \pi \cdot \dot{\vec{\pi}} \\
- c_s^2 [\partial \pi] \det \partial \pi - \frac{1}{8} c_s^2 [\partial \pi^2]^2 + \frac{1}{4} (c_s^2 + f_3) [\partial \pi^2] [\partial \pi]^2 - \frac{1}{24} (3 c_s^2 + 6 f_3 + f_4) [\partial \pi]^4 \\
+ \dot{\vec{\pi}} \cdot \partial \pi^2 \cdot \dot{\vec{\pi}} - (1 + c_s^2) [\partial \pi] \dot{\vec{\pi}} \cdot \partial \pi \cdot \dot{\vec{\pi}} + \frac{1}{2} [\partial \pi^T \cdot \dot{\vec{\pi}}]^2 \\
+ \frac{1}{4} ((1 + 3 c_s^2 + f_3) [\partial \pi]^2 - (1 + c_s^2) [\partial \pi^2]) \dot{\vec{\pi}}^2 + \frac{1}{8} (1 - c_s^2) \dot{\vec{\pi}}^4 \right\}. 
$$

where we have used $\partial \pi$ to denote the matrix with entries $(\partial \pi)_{ij} = \partial_i \pi_j$, and the brackets $[\ldots]$ to denote the trace of the matrix within. The first line collects the trilinear interactions, whereas the second and third lines collect the quartic ones. $f_3$ and $f_4$ stand for $f'''(1)$ and $f''''(1)$, respectively. Finally, notice that via the suffix $T$ we indicate the transpose of a matrix, rather than the transverse part of $\vec{\pi}$ as we did above.

### 2.1.5 Duality Between Irrotational Ordinary Fluid and Superfluid

In this subsection, we want to establish a classical duality between an irrotational (i.e. free of vortex modes) ordinary fluid with zero chemical potential (c.f. 2.6) and a zero-temperature superfluid.

The low energy description of a zero-temperature superfluid consists of one gapless scalar field $\psi$ — the Goldstone boson — which non-linearly realizes a spontaneously broken U(1) symmetry:

$$
\psi \rightarrow \psi + a.
$$

The effective Lagrangian for $\psi$ should then be the most general one compatible with the this shift-symmetry as well with Poincaré invariance, organized as a derivative expansion.

---

²For this reason in the rest of the thesis, we will not distinguish the spatial label “$i, j, \ldots$” from the internal ones “$I, J, \ldots$”.
At the lowest order in derivatives, it is given by

\[ \mathcal{L} = P(X), \quad \text{with} \quad X \equiv -\partial_\mu \psi \partial^\mu \psi. \]  

(2.29)

where \( P \) is a generic function, determining the equation of state. As was done in previous subsection, we can calculate the energy momentum tensor. We find that it also takes the form of (2.7), with the energy density, pressure and the velocity field given by

\[ \rho = 2XP' - P, \quad p = P, \quad u^\mu = \frac{\partial^\mu \psi}{\sqrt{X}}. \]  

(2.30)

The current associated with the \( U(1) \) shift-symmetry (2.28) is

\[ j^\mu = 2P'(X)\partial^\mu \psi. \]  

(2.31)

The ground state for the superfluid system is specified, in some proper units, by

\[ \psi = t, \]  

(2.32)

which agrees with the intuition that the superfluid ground state has finite charge density but vanishing spacial current.

On the other hand, the zero-chemical-potential ordinary fluid system, in which vortex excitations are absent, also has only one scalar degree of freedom — i.e. the longitudinal sound wave mode (c.f. 2.25). Moreover, at the classical level, it is consistent to suppress the appearance of the vortex modes. This is because Kelvin’s theorem states that the vorticity of fluids is conserved along the flow — that is, if we demand that the initial state be free of vortex modes, there is no vortex generated afterwards.

In our effective field theoretical language, the vorticity tensor of fluids is defined as an exact two-form:

\[ \omega \equiv d \left( 2F'(B)\sqrt{B} u_\mu dx^\mu \right), \]  

(2.33)

and the vanishing vorticity condition thus is

\[ \omega = 0. \]  

(2.34)

We claim that under the condition (2.34), the effective actions of ordinary fluids (2.6) and that of superfluids (2.29) are classically equivalent. The variables and functional dependences of these two sectors are related by

\[ \begin{cases} B = X(2P'(X))^2, & X = B(2F'(B))^2, \\ P(X) - F(B) = -2BF'(B) = 2XG'(X). \end{cases} \]  

(2.35)
The most straightforward way to understand this equivalence is that the energy momentum tensors of the two sectors are equal when the identification (2.35a) (2.35b) is imposed. There is a more elegant method to illustrate this duality relation and there are many subsequent computations one can perform related to this duality, we will refer interested readers to [1,32] for more details.

2.2 EFT for Imperfect Fluids

2.2.1 The general idea

Clearly, a field theory with a local action is non-dissipative by construction. But so is Nature: In any physical system, we call ‘dissipation’ the transfer of energy from the degrees of freedom we are interested in (collectively denoted by \( \phi \), in the following) to others which we are not keeping track of (collectively denoted by \( \chi \)), either because we are not concerned about them, or because describing them is too complicated or impractical. So, the best way to approach dissipation from a field theory viewpoint—at least conceptually—is to keep in mind that these additional degrees of freedom should also appear in the action of the system. That is, if we were to write the full action for \( \phi \) and \( \chi \), we would have

\[
S[\phi, \chi] = S_0[\phi] + S_\chi[\chi] + S_{\text{int}}[\phi, \chi].
\]  

(2.36)

\( S_0 \) is the action we would write for \( \phi \) alone, if we forgot about \( \chi \). \( S_\chi \) governs the dynamics of \( \chi \). \( S_{\text{int}} \) couples the two sectors, and is responsible for exchanging energy between them. If we now compute observables involving our \( \phi \) only, we can detect ‘dissipative’ effects—corresponding to exciting the \( \chi \) degrees of freedom—which cannot be reproduced by using \( S_0 \) alone. For instance, the \( S \)-matrix restricted to the \( \phi \)-sector is non-unitary whenever producing \( \chi \)-excitations is energetically allowed.

---

3The formal trick of adding an explicit time-dependence to a Lagrangian to make the energy not conserved—see e.g. [33]—might work to reproduce the desired dissipative equations of motion, but (i) is not systematic, i.e. it is not clear what the rules of the game are, and has therefore no predictive power, and more importantly (ii) does not correspond to the physical origin of dissipation, which is that there are additional degrees of freedom that have been ignored.
In the particular case we are interested in, \( \chi \) stands for the degrees of freedom of the microscopic constituents making up the fluid. For instance, for a weakly coupled, non-relativistic fluid made up of massive point-particles, \( \chi \) stands for the positions of these particles. On the other hand, \( \phi \) stands for the collective degrees of freedom, like sound waves for instance, which are those explicitly kept by the hydrodynamical description. Notice that at all times we are dealing with one and the same fluid, and its microscopic constituents. The splitting in Eq. (2.36) is one of the key features of the EFT formalism and the (emergent) dynamics in the long-wavelength limit. Hydrodynamics is about the dynamics of \( \phi \).

To illustrate the general idea, in this subsection we will not commit to the hydrodynamical case, nor will we go into many details. Rather, we will keep the discussion as general and as schematic as possible. We will only assume that the interaction Lagrangian \( S_{\text{int}} \) can be treated as a small perturbation. If this is not the case—if the two sectors are strongly coupled to each other—then it is not even clear how to talk separately of the \( \phi \)-sector and of the \( \chi \)-sector. In other words, we are assuming that as a first approximation, one can neglect the \( \chi \)'s when talking about the \( \phi \)'s. For hydrodynamics, as will see, this will be guaranteed by the symmetries: at low frequencies and momenta, all the interactions of the \( \phi \)'s become negligible, including those with the \( \chi \)'s. Notice that we are not assuming anything about interactions within the \( \chi \) sector: they can be arbitrarily strong.

Now, the crucial question is how to make use of expression (2.36), without actually specifying what the \( \chi \)'s and their dynamics really are. We will apply a method similar to that in Ref. [34]. The idea is to make the dependence of the interaction piece \( S_{\text{int}} \) on \( \phi \) explicit, while keeping that on \( \chi \) implicit. Schematically:

\[
S_{\text{int}} = \int d^4 x \sum_{n,m} \partial^n \phi^m(x) \mathcal{O}_{n,m}(x) .
\]

(2.37)

The \( \mathcal{O} \)'s are ‘composite operators’ of the \( \chi \)-sector—local combinations of the \( \chi \)'s and their derivatives. As usual, one expects all couplings allowed by symmetry to appear in the action. So, in particular, the \( \mathcal{O} \)'s should carry spacetime and possibly internal indices in

---

4 Strictly speaking, to avoid double counting, one should remove from the \( \chi \)'s the combinations of the individual particle positions that make up the \( \phi \)'s.
order to make the combinations appearing in $S_{\text{int}}$ invariant under all the symmetries that act on the $\phi$’s. Apart from symmetry, as usual in EFT, the other organizational principle in the infinite series \[ (2.37) \] is the derivative expansion: terms with fewer derivatives acting on the long distance/low energy degrees of freedom ($\phi$) matter the most at low energies and momenta.

Now, in any observable that involves measuring the $\phi$’s only—like for instance a $\langle \phi \phi \cdots \phi \rangle$ correlation function, or a $\phi \phi \rightarrow \phi \phi$ scattering amplitude—all effects due to the presence of the $\chi$’s, dissipative or otherwise, are “mediated” by the correlation functions of these $\mathcal{O}$ composite operators. As an example, consider a coupling (linear in $\phi$) between the two sectors of the form

\[ S_{\text{int}} = \lambda \int d^4 x \phi \mathcal{O} , \] \[ (2.38) \]

where $\lambda$ is a small coupling constant. For instance, suppose that we are interested in computing the $T$-ordered two-point function of $\phi$ in the standard vacuum (i.e. the vacuum for both the $\phi$ sector and the $\chi$ sector, for as we will see in a moment, computing this same correlator with a non-vacuum dissipative $\chi$ sector will necessarily complicate the story). This two-point function will receive contributions from $S_0$ and from $S_{\text{int}}$. We can compute the latter contribution in perturbation theory for $\lambda$. For instance, if $\phi$’s only interaction is that contained in $S_{\text{int}}$ above, this would correspond to the Feynman diagram series of Fig. [2.1]. In that case, neglecting combinatoric factors, powers of $i$, and momentum-conserving delta-functions, we would have schematically

\[ \langle \phi(p)\phi(-p) \rangle = \langle \phi(p)\phi(-p) \rangle_0 \]
\[ + \lambda^2 \langle \phi(p)\phi(-p) \rangle_0^2 \langle \mathcal{O}(p)\mathcal{O}(-p) \rangle_0 \]
\[ + \lambda^4 \langle \phi(p)\phi(-p) \rangle_0^3 \langle \mathcal{O}(p)\mathcal{O}(-p) \rangle_0^2 + \cdots , \] \[ (2.39) \]

where $T$-ordering is understood, and the subscript zeroes denote that those two-point functions are to be computed at zeroth order in $\lambda$, that is, in the absence of any interactions between $\phi$ and $\chi$. Once $\langle \phi\phi \rangle_0$ and $\langle \mathcal{O}\mathcal{O} \rangle_0$ are known, the full $\langle \phi\phi \rangle$ can be computed at any order in $\lambda$, without any further explicit reference to the $\chi$ dynamics. This is analogous to the standard Feynman-diagram expansion for a perturbative QFT, which involves the free propagators only. Here the correlators on the r.h.s. are not the free ones—they are those.
Figure 2.1: *Feynman diagram representation of Eqn. (2.39): the solid lines represent the φ propagators and the gray circles the two-point function of O.*

determined by $S_0$ (for φ) and by $S_\chi$ (for O) separately. In a more general case, where φ has non-trivial self-interactions and couples to the O’s in a more general way, the right-hand side looks more complicated because it involves higher-point correlation functions of φ and O as well. However, all the correlators are still evaluated at zero coupling ($\lambda$) between the two sectors.

As hinted at before, this simple picture gets slightly more complicated for correlation functions in more general states and in particular, thermal states. As we will discuss at some length in the next section, we will be interested in a thermalized χ sector. Its real-time correlation functions and the associated perturbative expansion then have to be handled via the so-called In-In, or Schwinger-Keldysh, formalism (for extensive reviews, see e.g. [35–37]). This entails a doubling of the fields in the path-integral, φ → φ±, χ → χ±, which complicates somewhat the systematics of the Feynman-diagram expansion. However, for what we are interested in, we can instead consider the effective (linearized) equations of motion for the expectation value of φ that we get by “integrating-out” the χ sector via In-In path integrals, which is essentially an In-In generalization of the quantum effective action formalism that is appropriate for systems described by a density matrix.

Following the notation of [35] and utilizing the simple coupling given by (2.38), the In-In generating functional for the correlation functions of φ is given schematically by

$$e^{iW[J_+,J_-]} = \text{const} \times \int D\phi_\pm D\chi_\pm$$

$$e^{i(S_0[\phi_\pm]J_+ \phi_\pm + S_\chi[\chi_\pm]J_+ \chi_\pm + \lambda \phi_\pm O_\pm)} ,$$

where the functional integral over χ+ and χ− is understood to include a (thermal) density matrix $\rho(\chi_0^+, \chi_0^-)$ for the initial conditions, which are also integrated over [36-37]. As we will see in a second, we will not need to be explicit about this.

Let’s assume that $\langle O \rangle = 0$ and confine ourselves to quadratic order in the $\phi_\pm$ fields. Noticing that, from the viewpoint of the χ sector, the $\phi_\pm$ fields act as external sources for
the operators \( O_\pm \), we can formally perform the functional integration over \( \chi_+ \) and \( \chi_- \) and obtain

\[
e^{iW[J_+, J_-]} = \text{const} \times \int D\phi_\pm e^{i S_2[\phi_\pm] \pm J_\pm \phi_\pm} e^{i \frac{\lambda^2}{2} \phi^a G_{ab} \phi^b},
\]

(2.42)

where \( S_2 \) is the quadratic action for \( \phi, \phi^a \equiv (\phi_+, -\phi_-) \), and \( G_O \) is a matrix of \( OO \) correlators [35]:

\[
G_O(x_1, x_2) = \begin{pmatrix}
\langle TO(x_1)O(x_2) \rangle & \langle O(x_1)O(x_2) \rangle \\
\langle O(x_2)O(x_1) \rangle & \langle O(x_1)O(x_2)T \rangle
\end{pmatrix}
\]

(2.43)

(The \( T \) to the right of a sequence of operators implies anti-time ordering.) These correlators have to be understood as traces involving the density matrix that is appropriate for the \( \chi \) sector.

The In-In effective action \( \Gamma [\phi_+, \phi_-] \) is then just the Legendre transform of the In-In generating functional \( W[J^+, J^-] \), from which the effective equations of motion for \( \langle \phi \rangle \) follow simply as [35]

\[
\frac{\delta \Gamma}{\delta \phi_+(x)} \bigg|_{\phi_+ = \phi_- = \langle \phi \rangle} = 0,
\]

(2.44)

However, since we are working at quadratic order in \( \phi_\pm \), the effective action \( \Gamma \) is just whatever appears at the exponent in the path integral (2.42) after having set \( J_\pm \) to zero:

\[
\Gamma_2[\phi_+, \phi_-] = S_2[\phi_+] - S_2[\phi_-] + \frac{\lambda^2}{2} \phi^a G_{ab} \phi^b,
\]

(2.45)

where two convolutions are understood for the last term. We thus get that the linear equation of motion for the expectation value of \( \phi \)—which, to keep the notation light, we also call \( \phi \)—is simply

\[
\frac{\delta S_2}{\delta \phi} + i \lambda^2 \langle OO \rangle_R \ast \phi = 0,
\]

(2.46)

where the second term involves precisely the retarded two-point function of \( O \):

\[
\langle O(x_1)O(x_2) \rangle_R \equiv \theta(t_1 - t_2) \langle [O(x_1), O(x_2)] \rangle.
\]

(2.47)

Note that the above conforms to the expectations of the usual “linear-response theory” result. What’s nice about the In-In formalism is that it allows one to generalize such a result to all orders in perturbation theory in a systematic fashion.

Keeping these qualifications in mind, and coming back to the main message of this subsection: For generic \( S_{\text{int}} \), in order to compute observables that involve the \( \phi \)'s only—and
in particular the time-evolution of $\langle \phi(x) \rangle$—we need not be explicit about the dynamics of the $\chi$’s. We ‘only’ need the $n$-point correlation functions of the operators the $\phi$’s couple to. Of course, knowing all such correlators is essentially equivalent to having solved the theory defined by $S_\chi$, which, as we stressed, can be arbitrarily complicated, strongly coupled, or simply unknown. Fortunately, in our particular case of hydrodynamics, we are interested in such correlators at very low frequencies and very long distances only. Moreover, we can assume that the $\chi$ sector—whatever it is—is in a state of thermal equilibrium. As we will see, this allows us to parameterize the leading low-frequency, long-distance behavior of the relevant correlators by three coefficients only.

2.2.2 Low frequency, long distance behavior of correlators

Consider the two-point function for a generic operator in the $\chi$ sector, $\langle O(\vec{x}, t) O(\vec{x}', t') \rangle$. If, in the absence of external perturbations—due for instance to our $\phi$’s—, the $\chi$ sector is thermalized, then the average $\langle \ldots \rangle$ has to be interpreted as a thermal trace with a density matrix $\rho \propto e^{-\beta H}$, or, in the presence of a conserved charge, $\rho \propto e^{-\beta(H - \mu Q)}$. (We will use a quantum mechanical language, but everything we say applies straightforwardly to classical statistical systems as well.) Now, we will assume that we identified correctly all the degrees of freedom that can propagate at long distances and for long times—we called them $\phi$—and that we constructed the most general EFT for them, encoded by $S_0[\phi]$. We will be more explicit in the next section, but for the moment, it suffices to say that these $\phi$’s correspond to the degrees of freedom traditionally associated with hydrodynamics: long-wavelength fluctuations in the energy density, in the velocity field, in the charge density, etc. Following the traditional language, we have ‘hydrodynamic modes’—i.e. physical variables with non-trivial long-range, late time correlators—for each conserved quantity: energy, momentum, charge. It is usually believed that thermal equilibrium erases all other information that is not associated with conserved charges. In particular, it is usually believed that in a thermal system correlators for quantities that are not densities for conserved charges decay rapidly, faster than any power, at very large distances and at very late times—roughly speaking, at distances and times larger than the mean free path and the mean free time, respectively.

Following this intuition, we will assume that the $\chi$-sector only features such rapidly
decaying correlators. As we will see, this does not imply that it does not feature gapless excitations. Indeed: if there were no gapless $\chi$-excitations, it would not be possible for very low frequency $\phi$ fields to transfer any energy to the $\chi$ sector. That is: at frequencies lower than the gap, there would be no dissipation whatsoever. Now, if an $\langle OO \rangle$ correlator decays faster than any power at large space- and time-separations, then its Fourier transform

$$G(\omega, \vec{k}) \equiv \int d^3x dt e^{i(\omega t - \vec{k} \cdot \vec{x})} \langle O(x,t)O(0) \rangle ,$$

is differentiable for real $\omega$ and $\vec{k}$—infinitely many times—at $\omega = \vec{k} = 0$. In particular, it admits a Taylor expansion in powers of $\omega$ and $\vec{k}$ about the origin. This means that at very low frequencies and momenta, we can parameterize our two-point function by just a few numbers—the coefficients of the leading terms in such a Taylor expansion.

To develop some physical intuition, it is useful to rephrase the above statement in terms of the spectral density for the operator $O$. So far we have been cavalier about the ordering of operators inside the two-point function. As pointed out in the last section, we will be mostly interested in the retarded two-point function,

$$G_R(x', t) \equiv \theta(t) \langle [O(x', t), O(0)] \rangle ,$$

which describes the causal response of the system to external disturbances, in the sense that adding a term $\int d^3x J O$ to the Hamiltonian—where $J(x, t)$ is a given external source—triggers a response in the expectation value of $O$

$$\langle O(x, t) \rangle_J = -i \int_{-\infty}^{\infty} dt' d^3x' G_R(x - x', t - t') J(x', t') + O(J^2)$$

(we have assumed that the background expectation value of $O$ vanishes, i.e $\langle O(x, t) \rangle_{J=0} = 0$). Its Fourier transform admits the spectral representation

$$G_R(\omega, \vec{k}) = \int_{-\infty}^{+\infty} d\omega_0 \frac{i}{\pi} \frac{i}{\omega - \omega_0 + i\epsilon} \rho(\omega_0, \vec{k}) ,$$

where $\rho(\omega_0, \vec{k})$—the spectral density—is a real, non-negative function (for positive $\omega_0$) that quantifies the density of states the system has at energy $\omega_0$ and momentum $\vec{k}$, weighed by
CHAPTER 2. EFT CONSTRUCTION FOR FLUIDS

the overlap the operator $\mathcal{O}$ has with them.  

One is often interested in separating the real and imaginary parts of Fourier-space correlation functions, because they contribute to different phenomena. In particular, the dissipative effects we are after will be associated with the imaginary part of $iG_R$, which, given the distributional identity

$$\frac{1}{x + i\epsilon} = P \frac{1}{x} - i\pi \delta(x) ,$$

is simply the spectral density:

$$\text{Im}(iG_R(\omega, \vec{k})) = \rho(\omega, \vec{k}) .$$

Our discussion following (2.48) thus implies that the spectral density should be infinitely differentiable for real $\omega$ and $\vec{k}$ at $\omega = \vec{k} = 0$, and that it should admit a low-frequency, low-momentum Taylor expansion. Moreover, standard arguments (see e.g. [38]) imply that the imaginary part of $iG_R$ is odd under $\omega \to -\omega$ (while the real part is even), so that in the Taylor expansion of $\rho$ we only have odd powers of $\omega$. The dependence on $\vec{k}$ is constrained by rotational invariance. If $\mathcal{O}$ is a scalar operator, it has to involve $1, |\vec{k}|^2, |\vec{k}|^4, \ldots$; If $\mathcal{O}$ carries a vector index $i$, the $\vec{k}$-dependence of the tensor spectral density $\rho_{ij}$ will involve the combinations $\delta^{ij}, k^i k^j, |\vec{k}|^2 \delta^{ij}, \ldots$; And so on for higher rank tensors. Given these properties, at very low frequencies and momenta, the spectral density of a tensor operator that transforms irreducibly under rotations can be parameterized by just one number—the first coefficient in its Taylor expansion:

$$\rho(\omega, \vec{k}) \simeq A \omega \times \delta \cdots \delta , \quad \omega, k \to 0 ,$$

where $\delta \cdots \delta$ stands for the combination of Kronecker-deltas with the right symmetries. Notice that $A$ has to be positive, because $\rho$ is positive for positive $\omega$.

---

$^5$The finite-temperature spectral density is given by

$$\rho(\omega, \vec{k}) = \frac{1}{2} \left( 1 - e^{-\beta \omega} \right) \left( \text{Tr} e^{-\beta H} \right)^{-1} \sum_{n,m} e^{-\beta E_n}$$

$$\times (2\pi)^4 \delta(\omega + E_n - E_m) \delta^3(\vec{k} + \vec{p}_n - \vec{p}_m) \langle n | \mathcal{O}(0, \vec{0}) | m \rangle^2$$

from which the non-negativity (for positive $\omega$) follows immediately.

$^6$For any operator $\mathcal{O}$ of given spin $s$, there is only one possible such combination that can appear in the
We thus see that the absence of long-range, late-time correlations in the $\chi$ sector does not forbid the existence of gapless excitations. These can exist, as long as the zero-momentum density of states (i) is a regular continuum in a neighborhood of $\omega = 0$, and (ii) goes to zero at zero frequency, at least as fast as $\omega$. For instance, a $\delta$-function contribution to the spectral density, peaked at $\omega = 0$, is not allowed. This would correspond to a gapless `single particle’ pole in correlators—i.e. to an excitation with a power-law propagator at very long distances and at very late times. According to our assumptions above, this should be included in the $\phi$ sector.

2.2.3 Dissipative couplings: $S_{\text{int}}$ to linear order

In this subsection, we will work out the low energy effective action for a dissipative fluid, under the guidance of the general ideas spelled out in the previous subsection. In particular we will focus on a fluid with zero chemical potential to make the discussion more accessible. For the finite chemical potential fluids, the system becomes much more complicated and the discussion about dissipation involves some subtleties. We will not discuss the finite chemical potential fluids here; interested readers can find details in [20].

Not surprisingly, we will adopt the low energy effective action (2.6) as our $S_0$; now we only need to write down the couplings of our $\phi^I$’s to the $\chi$-sector. There is one physical property of the $\chi$’s that we have not yet been explicit about: in a sense that we will try to make precise, these degrees of freedom “live in the fluid”, simply because they “make up” the fluid—they are supposed to describe all the degrees of freedom of the fluid’s microscopic constituents that are not explicitly taken into account by the $\phi$’s. This requirement alone should fix their transformation properties under all the symmetries that act on the $\phi$’s.

$\langle O O \rangle$ correlator. The reason is that in the tensor product of two spin $s$ representations, the singlet (spin-0) representation appears only once:

\[(2s + 1) \otimes (2s + 1) = 1 \oplus 3 \oplus \cdots \oplus (4s + 1) .\]  \hspace{1cm} (2.56)

For instance, if $O_{ij}$ is symmetric and traceless, that is, spin 2, its two point function at zeroth order in $\vec{k}$ has to take the form

\[\langle O_{ij} O_{kl} \rangle \propto \delta^{ik} \delta^{jl} + \delta^{il} \delta^{jk} - \frac{2}{3} \delta^{ij} \delta^{kl} .\]  \hspace{1cm} (2.57)
In what follows we will restrict ourselves to the lowest order in the derivative expansion and, more importantly, to linear order in the $\pi$ fluctuations where, as we shall see shortly, the coupling to the $\chi$ sector can be read off from basic properties of Goldstone boson interactions. In order to generalize our results to higher orders in the Goldstone fields, we would need to apply systematically the so-called coset construction to our case. We will leave addressing such a systematic construction to future work.

Let us parametrize the fluid’s ground state and the perturbation in a slightly different way (just a change in the normalization) from previous subsections. Suppose we start from an equilibrium configuration in which our Goldstones $\pi^I$ are set zero,

$$\phi^I_0(x) = b_0^{1/3}x^I . \tag{2.58}$$

Then, let’s turn on a small $\pi^I$ perturbation,

$$\phi^I(x) = b_0^{1/3} \cdot (x^I + \pi^I(x)) , \tag{2.59}$$

with very mild spatial gradients and time-derivatives. Since $\pi^I$ appears as an addition to $x^I$, this is equivalent to performing a small spatial translation of the original equilibrium field configuration (2.58), weakly modulated in space and time:

$$\phi^I_0(\vec{x}) \to \phi^I(\vec{x}, t) = \phi^I_0(\vec{x} + \vec{\pi}(\vec{x}, t)) . \tag{2.60}$$

We can now be precise about the meaning of “living in the fluid” for the $\chi$ sector: if the comoving coordinates $\phi^I$ are subject to a weakly modulated spatial translation as in eq. (2.60), the $\chi$ degrees of freedom undergo the same spatial translation. But, following standard Nöther theorem-type logic, under a modulated spatial translation with parameter $\vec{\pi}(\vec{x}, t)$, the $\chi$ action changes by

$$S_\chi[\chi] \to S_\chi[\chi] - \int d^4x \partial_\mu \pi^i T^{\mu i}_\chi , \tag{2.61}$$

where $T^{\mu i}_\chi$ is, by definition, the $\chi$ sector’s contribution to the Nöther current associated with spatial translations, that is, the spatial columns of the $\chi$ sector’s stress-energy tensor. Therefore we conclude that, at linear order in $\pi^i$,

$$S_{\text{int}} \simeq -\int d^4x \partial_\mu \pi^i T^{\mu i}_\chi . \tag{2.62}$$
Note that $\partial_{\mu}T^{\mu\chi} \neq 0$ and so the above interaction is non-trivial, since we are not including in $T^{\mu\chi}$ the $\pi$-dependent pieces that are required for conservation of the total stress-energy tensor.

A couple of comments about this expression are in order. First, the coupling above, while invariant under spatial translations, rotations, and $\pi$-shifts, does not seem to respect the volume-preserving symmetry of eq. (2.4). At linear order this symmetry requires invariance under

\[ \pi(t, \vec{x}) \to \pi(t, \vec{x}) + \vec{\epsilon}(\vec{x}), \quad \nabla \cdot \vec{\epsilon} = 0 . \] (2.63)

Since the $\vec{\epsilon}$ parameters are time-independent, we note that the 0-component of eq. (2.62) does respect the symmetry, whereas the spatial parts do not. At the moment we have no satisfactory understanding of this issue, but we are confident that (2.62) describes the correct linearized coupling of $\pi$ to the $\chi$ sector, because, as we will see in the next section, it correctly reproduces the first-order dissipative effects of hydrodynamics.

Second, the linear coupling of a Goldstone boson to the associated current—which we motivated via our “living in the fluid” logic—is likely a very general feature of theories with spontaneously broken symmetries. In the Appendix A we show that the analog of our coupling holds for a generic theory with a spontaneously broken internal $U(1)$ symmetry, and the logic of that example suggests that analogous results should apply for more generic (internal) symmetry breaking patterns. For spontaneously broken spacetime symmetries there will be additional subtleties, but we content ourselves by simply ignoring them for the moment.

Nevertheless, we will now show that (2.62) reproduces correctly the first-order dissipative phenomena associated with bulk and shear viscosity—including the celebrated Kubo

\[ \langle 0|j_{\mu}|\vec{p}\rangle \neq 0 . \] (2.64)

This interpolation property implies that the full current has terms that are linear in the Goldstone field, e.g. for relativistic theories $j_{\mu} = f \partial_{\mu} \pi + \ldots$. Here instead we are focusing on the terms in the current that depend on other fields—our $\chi$’s—but not on $\pi$, and we are claiming that, in the Lagrangian, the linear coupling of $\pi$ to this other sector involves precisely this $\pi$-independent part of the current.
relations. This give us the confidence in correctness of our formalism, despite that there is still a lack of complete understanding.

### 2.2.4 Rediscovering Kubo relations

As advertised in Section [2.2.1](#), we can now compute observables that involve our Goldstone excitations, and the \(\chi\) sector will contribute indirectly to these observables only via the correlators of the composite operators that couple to our Goldstones. Since the only couplings that we have so far are linear in the Goldstones, the observables we are able to compute at this point have to do with the free propagation of Goldstone excitations. That is, we are able to compute the Goldstone attenuation rates.

Once again in this subsection we will only consider a fluid without conserved charges. Its excitation spectrum—neglecting dissipative effects—is described by the action (2.24). We have a longitudinal mode \(\pi_L\) with \(\omega = c_s k\), and two transverse modes \(\pi_T\) with a degenerate dispersion relation \(\omega = 0\). Consider now one such excitation propagating in the fluid. Its coupling to the \(\chi\) sector via the interaction (2.62) will make it slowly decay away, eventually transferring all its initial energy to \(\chi\) excitations. We can compute the rate at which this decay process takes place at the level of the classical equations of motion for the Goldstones. We could also do the computation at the level of Feynman-diagram perturbation theory, which would be more in line with our field theoretical approach. In particular, since the attenuation rates we are after correspond to imaginary shifts in the excitations’ frequencies, we should compute the \(\chi\)-mediated corrections to the poles of the \(\pi^I\) propagator. However, as reviewed in sect. [2.2.1](#) in the in-in formalism each propagator gets replaced by a \(2 \times 2\) matrix of propagators, which, at least for our simple computation, complicates unnecessarily the systematics of perturbation theory.

Following Section [2.2.1](#) the linearized eom for \(\pi^I\) derived from the Goldstone quadratic action (2.24), augmented by their interaction with the \(\chi\) sector (2.62), is precisely what one
would naively expect from linear response theory:

\[ w_0(\omega^2 \pi^i - c_s^2 k^i k^j \pi^j) + iG_R^{ij}(\omega, \vec{k}) \pi^i = 0, \quad (2.65) \]

where \( G_R^{ij} \) is the retarded two-point function of the combination that couples to \( \pi^i \) in (2.62):

\[ G_R^{ij}(\omega, \vec{k}) = k_\mu k_\nu \langle T_\mu^i \chi T_\nu^j \chi \rangle, \quad (2.66) \]

and from now on we will use simply \( \langle \ldots \rangle \) to denote the Fourier transforms of \textit{retarded} two-point functions, evaluated at \( \omega \) and \( \vec{k} \). Moreover, it will be understood that \( G_R \) is evaluated in Fourier space, and its \( \omega, \vec{k} \) arguments will be omitted.

In the end we are interested in the imaginary parts of the eigenfrequencies of the system, which—at leading order in perturbation theory—will be related to the imaginary part of \( iG_R \). At this point we could parameterize the infrared behavior of \( \text{Im} \cdot i\langle T_\mu^i T_\nu^j \chi \rangle \) as described in Section 2.2.2, but, before proceeding let us massage this quantity a little in order to rewrite it in a form that the reader familiar with hydrodynamics will recognize. First, notice that according to (2.54), such a quantity is the spectral density of a composite operator \( (T_\mu^i \chi) \) in the \( \chi \) sector. We argued that all local operators in the \( \chi \) sector should have very well behaved spectral densities near \( \omega = \vec{k} = 0 \), at least for real \( \omega \) and \( \vec{k} \), with a Taylor expansion starting as \( \text{const} \cdot \omega \), and continuing with higher powers of \( \omega \) and \( \vec{k} \). At low energies and momenta, we are interested in just that first term, which we can extract formally by taking the nested limit

\[ \omega \lim_{\omega \to 0} \left[ \frac{1}{\omega} \lim_{\vec{k} \to 0} \left( \text{Im} \cdot i\langle T_\mu^i T_\nu^j \chi \rangle \right) \right]. \quad (2.67) \]

Given the regularity of our spectral densities in the infrared, we can take the limits in any order. However, taking the limits in the order we have written them allows us to replace \( T_\chi^\mu \) with the \textit{total} \( T^\mu \), which includes contributions coming from the Goldstone bosons. The reason is that at lowest order in the \( \chi-\pi \) interactions and in the derivative expansion, the Goldstones’ contribution to any spectral density is a Dirac-delta peaked at on-shell values for

---

\(^8\)We have assumed, as we did in Section 2.2.1, that \( \langle T_\chi^\mu \rangle_{\pi=0} \) vanishes. In our formalism, the equilibrium expectation value for the fluid’s \textit{full} stress energy tensor is given by \( \langle \ldots \rangle \), evaluated at the equilibrium configuration \( \langle \ldots \rangle \), that is, it is fully captured by the \( \phi \)'s sector action \( S_0 \).
ω and ⃗k. But the limit in (2.67) carefully dodges such on-shell values, both for longitudinal 
(ω = c_s ⃗k) and for transverse (ω = 0) excitations. At the order we are working we thus have

\[ \text{Im}(iG_{ij}^{ij}) \simeq \omega k_{\mu} k_{\nu} \cdot \lim_{\omega \to 0} \frac{1}{\omega} \lim_{\vec{r} \to 0} \text{Im} i \cdot \langle T^{\mu i} T^{\nu j} \rangle \]

(2.68)

Then, using a standard trick—see e.g. [39]—we can use conservation of the full stress-energy tensor to set to zero terms in the correlator above that have μ or ν equal to zero. The reason is that, because of \( T_{\mu \nu} \) conservation, in Fourier space we have the operator equation

\[ T^{\alpha \alpha} = \frac{k^\alpha}{\omega} T^{k \alpha}, \]

(2.69)

which yields zero if we take ⃗k to zero first, like we are doing in the limit above.⁹

We are thus left with

\[ \text{Im}(iG_{R}^{ij}) \simeq \omega k_{k} k_{l} \cdot \lim_{\omega \to 0} \frac{1}{\omega} \lim_{\vec{r} \to 0} \text{Im} i \cdot \langle T^{k i} T^{l j} \rangle . \]

(2.70)

Following Section 2.2.2, we can now split the stress tensor operator \( T^{ij} \) into irreducible representations of the (unbroken) rotation group, spin 0 and spin 2,

\[ T^{ij} = T_{0}^{ij} + T_{2}^{ij} \]

(2.71)

\[ T_{0}^{ij} = \frac{1}{3} \delta^{ij} T^{kk}, \quad T_{2}^{ij} = T^{ij} - \frac{1}{3} \delta^{ij} T^{kk}, \]

(2.72)

and parameterize the low-energy behavior of the associated spectral densities—in the nested limit we are interested in—via two free parameters \( A_{0,2} \) as

\[ \text{Im} i \cdot \langle T_{0}^{ki} T_{0}^{lj} \rangle \simeq A_{0} \omega \cdot \delta^{ki} \delta^{lj} \]

(2.73)

\[ \text{Im} i \cdot \langle T_{2}^{ki} T_{2}^{lj} \rangle \simeq A_{2} \omega \cdot \left( \delta^{kl} \delta^{ij} + \delta^{kj} \delta^{il} - \frac{2}{3} \delta^{ki} \delta^{lj} \right) . \]

We should also mention that the mixed correlator \( \langle T_{0} T_{2} \rangle \) vanishes at zero momentum, because of rotational invariance.

⁹The manipulations we just performed may seem dangerous: in fact, in the last section we insisted that is important that the \( \pi^i \) does not couple to the full \( T^{\mu i} \), but only to a non-conserved part of it, so that the coupling (2.62) is actually non-trivial. There is no contradiction however: the divergence—or the \( k^\mu \)—one needs to annihilate the full stress-energy tensor does not commute with our nested limit, so that the r.h.s. in eq. (2.68) is actually non-zero.
Plugging these parameterizations into eq. (2.70) we get

$$\text{Im}(iG_R^{ij}) \simeq \omega k^2 [(A_0 + \frac{4}{3} A_2) P_L^{ij} + A_2 P_T^{ij}] ,$$

(2.74)

where $P_{L,T}^{ij}$ are the longitudinal and transverse projectors

$$P_L^{ij} = \hat{k}^i \hat{k}^j , \quad P_T^{ij} = \delta^{ij} - \hat{k}^i \hat{k}^j .$$

(2.75)

The reason it’s convenient to split this contribution to the $\pi^i$ eom as a sum of a longitudinal and a transverse part, is that the zeroth-order eom has a similar structure:

$$\omega^2 \pi^i - c_s^2 k^i k^j \pi^j \to \left[ (\omega^2 - c_s^2 k^2) P_L^{ij} + \omega^2 P_T^{ij} \right] \pi^j .$$

(2.76)

Then, putting everything back into eq. (2.65) we get immediately the imaginary parts of the (low-momentum) eigenfrequencies:

$$\Delta \omega_L \simeq -i \frac{(A_0 + \frac{4}{3} A_2)}{2w_0} k^2$$

(2.77)

$$\Delta \omega_T \simeq -i \frac{A_2}{w_0} k^2 .$$

(2.78)

These are the attenuation rates for, respectively, the longitudinal and transverse modes. We already see two important predictions of our field theoretical approach. First, the dissipative nature of the coupling (2.62): these imaginary frequency shifts have the right sign to make the Goldstone excitations decay in time, since the positivity of $A_{0,2}$ is guaranteed by the positivity properties of any spectral density, as reviewed in Section 2.2.2. Second, the attenuation rates scale as $k^2$ at low momenta, which agrees with the standard dissipative hydrodynamics results.

But we can go further. Comparing our attenuation rates to the standard ones in the literature—see e.g. [39,40]—we find that our parameters $A_{0,2}$ correspond to bulk and shear viscosity, usually denoted by $\zeta$ and $\eta$:

$$\zeta = A_0 , \quad \eta = A_2 .$$

(2.79)

Then, our definitions of $A_{0,2}$ in eq. (2.73), match precisely the famous Kubo relations for bulk and shear viscosity [39]. This is one of the main results of this thesis: an independent derivation of the Kubo relations via effective field theory techniques.
2.3 A Glance at the Quantum Mechanics of Perfect Fluids

2.3.1 A Naive Deformation

Now let us go back to perfect fluids with no conserved charge. The question we want to understand in this section is whether I can make use of the effective action (2.23) beyond the classical level and how to make sense of it in order that it can be treated as a quantum theory.

The structure of the quadratic Lagrangian (2.24) already signals that, upon canonical quantization, we might be facing a strong-coupling problem for the vortices. The reason is the following: Consider first as a toy model a quantum-mechanical oscillator with some anharmonic corrections to the potential. In perturbation theory, one first solves the harmonic problem, thus getting the standard oscillator spectrum, and then treats the anharmonicities as small corrections. The approximation is justified for those states whose wavefunctions are localized in a region where the potential is dominated by its quadratic approximation. So, for perturbation theory to be applicable in this case, one needs at least the ground state to have a localized enough wave-function (highly excited states will always be outside the regime of validity of perturbation theory in this example.) Of course, what localizes the ground state is the curvature of the harmonic potential—the oscillator’s frequency. For the system to be ‘weakly coupled’, one thus needs a deep enough quadratic potential. If we now move on to field theory, the role of the quadratic potential is usually played—in the absence of mass terms—by the gradient energy. For given spatial momentum $\vec{k}$, the gradient energy gives a potential $\propto k^2|\varphi_k|^2$. The vacuum wavefunction is thus localized about $\varphi_k = 0$, and cannot probe large field values where interactions may become important. In the absence of a gradient energy, on the other hand, each mode’s vacuum wavefunction is totally delocalized in the quadratic approximation, and its dynamics are completely determined by the interactions. We thus reach the conclusion that a (massless) field theory without gradient energies is prone to strong coupling, at all scales.

There is a number of caveats in applying the above logic to our case. The first is that the absence of gradient energy may be an accidental feature of the lowest order in the derivative expansion. This is the case, for instance, for the ghost condensate [41], where...
gradient energy starts at the four-derivative order, \( E_{\text{grad}} \propto (\nabla^2 \pi)^2 \). In the absence of quadratic terms with fewer spatial derivatives, such a term cannot be relegated to the class of higher-dimension operators, because it is marginal by definition—together with the kinetic energy \( E_{\text{kin}} \propto \dot{\pi}^2 \) it determines how things behave under rescalings. In this case then, there is a well defined perturbative expansion. But this way out is not available to our vortices: the absence of gradient energy for them is enforced by a symmetry, which also forbids higher spatial-derivative quadratic terms. In the absence of time-dependence, exciting vortices costs nothing: we can deform the ground state \( \phi^I = x^I \) in the ‘transverse’ direction via eq. (2.4) and pay no energy price, and this extends to non-linear order as well. The second caveat, more relevant for us, is that the above quantum oscillator toy model assumes that the anharmonic interactions are of the potential form—only in this case delocalization of the wavefunction necessarily leads to strong coupling, because having access to large values of \( q \) entails having access to large interactions. But in our case, by construction, we only have derivative interactions, and moreover the very same symmetry that forbids the vortex gradient energy is also going to forbid many interactions involving vortices. In particular, as we will see more concretely in the following, all vortex interactions that do not involve at least two time derivatives are forbidden. Therefore the connection between wavefunction delocalization and strong-coupling is less obvious in our case.

To settle the question, we should probe the theory by computing some physical quantity and check whether the perturbative expansion holds. The ideal candidates are usually \( S \)-matrix elements, but here we face a complication. The longitudinal phonon has standard wave solutions, which upon canonical quantization, get mapped onto standard free-particle states. The transverse phonons, in contrast, do not behave as waves, and as a consequence there are no quantum asymptotic states associated with them. The classical field \( \vec{\pi}_T \) behaves like a collection of infinitely many free particles rather than infinitely many oscillators. Upon quantization, its Hilbert space is not made up of standard Fock states. Without asymptotic states there is no \( S \)-matrix.

A possible alternative, is to compute instead local \( n \)-point functions in real space, and to check whether perturbation theory holds for them. They may be as physical as the \( S \)-matrix: they characterize the physical interaction among local sources that couple to our
fluid. We do not need asymptotic states to set up such a question. For instance, we can
define the theory and the associated correlation functions via the path-integral formulation.
Another possibility, which we will choose, is to give the theory asymptotic states for the
vortex degrees of freedom by deforming it in the IR. We can add to the classical action
a term that is compatible with all the symmetries except for the volume-preserving diffs
(Eqn. (2.4)),
\[ \Delta S = -w_0 \int d^4x \frac{1}{2} c_T^2 B^{II}, \quad c_T^2 \ll c_s^2 \] (2.80)
and whose only effect, once expanded about the ground state, is to introduce a small gradient
energy for the transverse Goldstones\[10\]
\[ S_2 \to w_0 \int d^4x \left[ \frac{1}{2} \dot{\pi}^2 - \frac{1}{2} c_s^2 \left( \nabla \cdot \pi_L \right)^2 - \frac{1}{2} c_T^2 \left( \nabla_i \pi^j_T \nabla_i \pi^j_T \right) \right]. \] (2.82)

2.3.2 Scattering Amplitude of Some Simple Processes — the Vortex
Strong Coupling

In this subsection, we will do some sample calculations of 2 to 2 scattering processes to
illustrate explicitly the peculiar quantum mechanical features of our theory associated with
vortex degree of freedom. For simplicity we will only consider initial states with zero
total momentum: given that Lorentz boosts are spontaneously broken and that we have
a preferred reference frame, this is a non-trivial choice—we are setting some kinematic
invariants to zero. With an abuse of language, we will refer to this choice as “working in
the center of mass (CM) frame.”

We will not content ourselves with amplitudes. Rather, we will compute physical, mea-
surable quantities like cross sections and decay rates. The reason is that amplitudes depend
crucially on the normalization chosen for the single-particle states. For instance going from
the so-called relativistic normalization to the non-relativistic one, would move some factors

\[ ^{10} \text{More precisely, the expansion of } B^{II} \text{ is} \]
\[ B^{II} = -\dot{\pi}^2 + 2 \nabla \cdot \pi + (\nabla_i \pi^j_T \nabla_i \pi^j_T ). \] (2.81)
The linear term is a total derivative, and can thus be neglected. The other terms, on top of giving
the transverse phonons a gradient energy, correct the kinetic and gradient energies already present in (2.24).
However in the limit \( c_T^2 \ll c_s^2 < 1 \) these corrections are also negligible.
of $c_s$ and $c_T$ from the amplitudes to the phase-space elements, in such a way as to keep cross-sections and rates unaffected. Ascertaining the strong-coupling of the theory in the $c_T \to 0$ limit at the level of amplitudes requires a derivation of partial waves, a la Jacob-Wick, being careful about the factors of $c_s$ and $c_T$. Thus we found it simpler to present them by focusing on cross-sections and decay rates.

A final remark about external vortices in initial states. When we take the $c_T \to 0$ limit we have to decide whether we are going to be keep their momenta or their energies fixed. The first choice is the more conservative, since it corresponds to taking their energies to zero, thus weakening any possible strong-coupling phenomenon we are going to encounter. It is also the only consistent one, since the alternative one would send the vortex momenta to infinity, outside the regime of validity of any effective theory. Notice also that only if we keep the vortex momenta fixed is our deformed theory with small $c_T$ close to the fluid one with $c_T = 0$: in the Lagrangian $c_T^2$ weighs the gradient energy, so that by sending $c_T$ to zero while keeping the momenta fixed one is in fact sending the magnitude of that Lagrangian term to zero.

We will use $\vec{p}$’s to denote the momenta of the longitudinal modes, and $\vec{k}$’s and $\hat{\epsilon}$’s to denote the momenta and polarizations of the transverse modes. Our $\hat{\epsilon}$’s are real, thus corresponding to linear polarizations, and normalized to one (hence the ‘hat’.) For all the processes we will just compute the leading contribution in the limit $c_T/c_s \ll 1$, for which we hope to learn something about the original fluid ($c_T = 0$).

**Longitudinal 2 → 2 Scattering**

This is the simplest of the scattering processes. To tree level, the only relevant diagrams are shown in Fig. 2.2. We designate, here and for the rest of the section, the solid lines as longitudinal excitations and the curly lines as the transverse excitations. Time flows to the right.

When done in the center of mass frame, the only kinematic variables are the momentum of the longitudinal phonons $p$ and the scattering angle $\theta$. To tree level, the total amplitude is given by

$$iM_{LL \rightarrow LL} = \frac{ip^4c_T^2}{u_0} \left\{ -3c_s^2(1-c_s^2) + 2(1-c_s^2)\cos^2\theta - 2f_3 + 2f_3^2/c_s^4 - f_4/c_s^2 \right\}$$

(2.83)
The infinitesimal cross section is

\[ d\sigma = \frac{1}{c_s^6} \frac{|\mathcal{M}_{LL\rightarrow LL}|^2}{64\pi^2(2p)^2} d\Omega, \quad (2.84) \]

where we made use of the phase space element computed in the Appendix B (eq. (B.7)). We can easily calculate the total cross section. The final particles are identical, so we over-count when we integrate over all final phase space. To counteract this we simply include a \(1/2\) symmetry factor. To all orders in \(c_s\) the total cross-section is

\[ \sigma_{LL\rightarrow LL} = \frac{1}{256\pi p^2} \left( \frac{p^4}{w_0 c_s} \right)^2 \left[ 2\alpha^2 - \frac{4\alpha\beta}{3} + \frac{2\beta^2}{5} \right] \sim \frac{1}{p^2} \left( \frac{p^4}{w_0 c_s} \right)^2 \quad (2.85) \]

where \(\alpha \equiv (f_4/c_s^2 - 2f_3^2/c_s^4 + 3c_s^2 + 2f_3 - 3c_s^4) = \mathcal{O}(1) + \mathcal{O}(c_s^2) + \mathcal{O}(c_s^4)\) (assuming \(f_3, f_4 \sim c_s^2\)) and \(\beta \equiv 2(1 - c_s^2)\).

From this result, we can read off the strong coupling momentum and energy associated with longitudinal phonons:

\[ p_{*L} = (w_0 c_s)^{1/4}, \quad E_{*L} = c_s p_{*L}. \quad (2.86) \]

The reason is the following: the cross-section (2.85) can be understood as the geometric cross-sectional area for wave-packets of wavelength \(1/p\), times the square of the dimensionless interaction strength. We have strong coupling when \(\sigma\) becomes of order \(1/p^2\)—in such a case the two wave-packets have an \(\mathcal{O}(1)\) probability of not missing each other.

**Transverse 2 \rightarrow 2 Scattering**

Now we turn to compute the cross section of the TT\(\rightarrow\)TT scattering process. The relevant tree-level Feynman diagrams are shown in Fig. 2.3. We find that there are dramatic
cancellations and the zeroth and first order terms in the limit $c_T/c_s \to 0$ in the amplitude $\mathcal{M}_{TT\to TT}$ vanish but, nevertheless, to leading order it is quadratically dependent on $c_T$.

More concretely, in the CM frame with $\parallel$ and $\perp$ denoting polarizations parallel and perpendicular to the scattering plane, respectively, at lowest order in $c_T$ the amplitude is

$$i\mathcal{M}_{TT\to TT} = \frac{i k^4 c_T^2}{w_0} \times \begin{cases} \cos 2\theta & \text{for } \parallel\parallel\to \perp\perp, \perp\perp\to \parallel\parallel \\ \frac{1}{2} (\cos \theta - \cos 2\theta) & \text{for } \parallel\perp\to \parallel\perp \end{cases}$$

(2.87)

and zero for all other combinations of polarizations. In the $\parallel\perp\to \parallel\perp$ case, $\theta$ is the angle between the two $\parallel$-polarized phonons. Note that there is no dependence on $f_3$ and $f_4$. So this process is generic for all fluid models regardless of the details of the equation of state — i.e. the functional dependence $f(\sqrt{B})$.

After squaring the amplitude, we average over incoming polarizations and sum over the final ones. We have:

$$\frac{1}{4} \sum_{\text{initial } \epsilon} \sum_{\text{final } \epsilon} |\mathcal{M}|^2 = \frac{1}{4} \left( \frac{k^4 c_T^2}{w_0} \right)^2 \left[ 2 + \frac{1}{2} \cos 2\theta + \frac{3}{2} \cos 4\theta \right] ,$$

(2.88)

and so it follows immediately from (B.7) that the total cross section, including a 1/2 symmetry factor, is

$$\sigma_{TT\to TT} = \frac{1}{256\pi} \left( \frac{13}{15} \right) \frac{1}{k^2} \left( \frac{k^4}{w_0 c_T} \right)^2 \sim \frac{1}{k^2} \left( \frac{k^4}{w_0 c_T} \right)^2 .$$

(2.89)

Thus the strong coupling momentum associated with vortex excitations (transverse modes) is

$$k_{*,T} = (w_0 c_T)^{1/4}, \quad E_{*,T} = c_T k_{*,T} .$$

(2.90)
The cross section blows up as we take $c_T/c_s \to 0$, signifying our theory becomes ill-defined at arbitrary external momenta in this limit. We emphasize once again the absence of the free parameters $f_3$ and $f_4$ in this result, which implies that the pathology just unveiled cannot be avoided by a judicious choice of their values.

2.3.3 The infrared situation

Our $S$-matrix analysis indicates that the transverse degrees of freedom are strongly-coupled at arbitrarily low energies. However the strong-coupling phenomenon we unveiled is quite peculiar. We deformed the theory in the far IR by introducing a small deformation parameter $c_T$. This changes the asymptotic states of the theory, and we discovered that some of these get strongly coupled in the UV, at an energy scale that drops to zero when we recover the original theory—the limit $c_T \to 0$. A vanishing ultraviolet strong-coupling energy scale suggests that our problem is probably more properly thought of as an infrared one—we may be approaching strong-coupling from the wrong side! That is, in the deformed theory at finite $c_T$ we encounter strong coupling in moving to high energies, but since the deformed theory differs from the original one at low energies, it may be that the strong coupling scale is in fact a divide between the two theories—that there is no regime where the two theories look alike. If we stick to the original description, without ever introducing $c_T$, we may realize that we have some form of strong coupling in the IR.

The distinction we are putting forward may sound like a matter of definition, but it is not. A theory that becomes strongly coupled in the UV is simply not defined at energies of order of the strong-coupling scale and above—it needs infinitely many parameters for its definition. If this were the case for us, our theory would not be consistent, in any energy range. On the other hand, there are a number of ways in which perturbation theory can break down in the IR without impairing the consistency of a theory. There is for instance real QCD-like strong coupling, where perturbation theory does break down but the non-perturbative theory is perfectly well defined—it is just hard to solve! Or there are QED-like infrared divergences, which can be tamed by focusing on suitable infrared-safe observables, for which the perturbative expansion applies. Or there may be huge quantum IR fluctuations without necessarily implying large interactions, like for instance for would-
be Goldstone bosons in 1 + 1 dimensions \cite{42}. This would signal a bad identification of the theory’s vacuum state. And in general, to ascertain the consistency of an IR-problematic theory, at least in certain energy and momentum regimes, one can just put the system in a finite-size ‘box’ and consider its time evolution for a short time, in which case perturbation theory typically does not even break down.

In the rest of this subsection, we will present a rough argument to demonstrate that our strong-coupling problems with the transverse degrees of freedom may be infrared in nature and stem from an incorrect definition of the vacuum, in an analogy to Coleman’s theorem \cite{42} in 1 + 1 dimensions. In other words, we have been assuming all along that doing perturbation theory about the semiclassical vacuum \( \phi^I = x^I \) is sensible, but somehow quantum fluctuations want to dismantle this state. (c.f. Eqn. (2.93) below) Interested readers are referred to \cite{2} for a more thorough discussion.

To start, let us consider quantum fluctuations about the semiclassical vacuum state with \( \langle \phi^I \rangle = x^I \). An order parameter that conveniently quantifies the amount of spontaneous symmetry breaking in a manifestly translationally invariant fashion is

\[
\langle \partial_\mu \phi^I \rangle = \delta^I_\mu \sim 1. \tag{2.91}
\]

It is straightforward to estimate quantum fluctuations in this quantity. We decompose the fields as \( \phi^I = x^I + \pi^I \), and from the \( \pi^I \) propagators,

\[
\langle T \pi^I(x) \pi^J(y) \rangle \rightarrow \frac{1}{w_0} \cdot \frac{iP^I_J}{\omega^2 - c_l^2 p^2 + i\epsilon} + \frac{1}{w_0} \cdot \frac{iP^I_T}{\omega^2 + i\epsilon}, \tag{2.92}
\]

where \( P_L \) and \( P_T \) are the longitudinal and transverse projectors, we get

\[
\lim_{\omega \to 0} \langle \partial_\mu \pi^I \partial_\nu \pi^J \rangle \sim \frac{PP^I_J}{w_0} \frac{p^5}{\omega} + \frac{PP^I_J}{w_0} \frac{p^3}{\omega} \tag{2.93}
\]

\[
\lim_{\omega \to 0} \langle \partial_0 \pi^I \partial_0 \pi^J \rangle \sim \frac{1}{w_0} p^4 \tag{2.94}
\]

\[
\lim_{\omega \to 0} \langle \partial_0 \pi^I \partial_0 \pi^J \rangle \sim \frac{1}{w_0} p^3 \omega \tag{2.95}
\]

These are real-space correlators, and in the right-hand sides the dimensionless, order-one part of the Fourier transform, \( \int d\Omega \, d\log p \, d\log \omega \, e^{i\omega \rho} \), is understood. Also we are considering considerably off-shell (\( \omega, p \)) pairs, by taking for instance the separation in real space to be space-like with the respect to the sound speed (i.e., by working in Euclidean space.) The
correlators (2.94, 2.95) behave in essentially the same way as for standard field theories in dimensions higher than 1 + 1: quantum fluctuations in order parameters are damped at low momenta and low energies, and as a consequence in the IR there is spontaneous symmetry breaking. However the correlator (2.93) ruins this familiar picture: for fixed momentum, it is IR-divergent. Equivalently, in the $p$-$\omega$ plane there is a sector extending all the way to $p = 0$, $\omega = 0$ where quantum fluctuations in our order parameter (2.91) are huge. And we also see that this peculiar feature is precisely due to the transverse excitations.

2.4 Application of the Effective Fluid Model to Cosmology

2.4.1 Perfect Fluids in a Cosmological Background

In cosmology, the matter content of the universe is often modeled as a perfect fluid. In our field theoretic language, it is straightforward to extend the effective field theoretic description of perfect fluids to include dynamical gravity effects. We write the action of the cosmological model as

$$S = S_{EH} + S_{\text{Fluid}}$$

(2.96)

where the first term on the r.h.s. is just the usual Einstein - Hilbert action for gravity and the matter action $S_{\text{Fluid}}$ is given by (2.6), with the flat metric $\eta$ replaced by a cosmological spacetime one $g$ and the measure by $\sqrt{-g} \, d^4 x$. For the sake of this section — i.e. to look for a conserved curvature perturbation $\zeta$ of the universe to the non-linear level — we should focus on fluids with no conserved charge [43].

Suppose that the metric fluctuates around a flat FRW background

$$\bar{g}_{\mu\nu} = \text{diag}\{-1, a(t)^2, a(t)^2, a(t)^2\}.$$  

(2.97)

The number of dynamical degrees of freedom in question is counted as follows. In addition to the three Goldstone fields $\pi^I$ of the fluid sector — which, as argued in Section (2.1), are eventually regrouped into one longitudinal scalar and one transverse vector under the

\footnote{Roughly speaking, the reason for this is that, in a system in thermal equilibrium where all conserved quantum numbers vanish, the existence of adiabatic $\zeta$ is guaranteed.}
residual SO(3) group — , the gravity sector introduces an extra d.o.f. — the spin−2 graviton, a traceless transverse tensor field.

Perhaps a more transparent way to see this is by removing the gauge redundancy of the gravity sector. Using the ADM variables, we know that only the spatial part of the metric is dynamical, which in complete generality can be decomposed as

\[ g_{ij} \equiv h_{ij} = a(t)^2 \exp (A \delta_{ij} + \partial_i \partial_j \chi + \partial_i C_j + \partial_j C_i + D_{ij}) \]  

(2.98)

where \( C_i, D_{ij} \) satisfy

\[ \partial_i C_i = 0, \quad \partial_i D_{ij} = D_{ii} = 0. \]  

(2.99)

If the four gauge conditions of coordinate transformations are chosen to set \( A = 0, \chi = 0, C_i = 0 \) — which is known as the spatially flat slicing gauge (SFSG), — we are left with \( \pi_L, \pi_T, \) and \( D_{ij} \) as the dynamical degrees of freedom, which characterizes, respectively, the scalar, vector and tensor cosmological perturbation.

For the interest of this section, it turns out to be most convenient to work in another gauge — called the unitary gauge (UG), — in which all the perturbations are absorbed into the metric, leaving the matter fields unperturbed: \( \phi^I = x^I \). Said differently, the spatial coordinates are chosen to coincide with the comoving of the fluid. Meanwhile the temporal gauge freedom is used to determine the time slices such that the scalar quantity \( B \) (defined in eqn. (2.5)) remains unperturbed on each time slice: \( B(t) = a(t)^{-6} \).

The unitary gauge leads to many conceptual and computational simplifications. First of all, with our choice of the spatial coordinates, the worldlines of fluid volume elements coincide with the threads \( x^i = \text{constant} \). Indeed, this can be seen by considering the spatial components of the velocity field, which vanish since \( u^I \propto \epsilon^{Iabc} \partial_a \phi^1 \partial_b \phi^2 \partial_c \phi^3 = 0 \). Second, our choice of time slices coincides with the uniform density slices, for the energy density of the fluid, given by \( \rho = -F(B) \), is only a function of \( B \) and it is a constant on each time slice.

Using the ADM variables, we can parametrize the metric as

\[ ds^2 = -N^2 dt^2 + h_{ij}(dx^i + N^i dt)(dx^j + N^j dt) \]  

(2.100)

and the inverse metric \( g^{\mu\nu} \) as

\[ g^{00} = -\frac{1}{N^2}, \quad g^{0i} = g^{i0} = \frac{N^i}{N^2}, \quad g^{ij} = h^{ij} - \frac{N^i N^j}{N^2} \]  

(2.101)
where the spatial metric \( h_{ij} \) takes the form of (2.98) and \( h^{ij} \) is the inverse spatial metric: 
\[ h^{ik} h_{kj} = \delta^i_j. \]
One can show easily that in the UG, \( B^{IJ} = g^{IJ} \) \(^{12}\) and hence that
\[
B = \det g^{IJ} = \frac{1}{\det h_{IJ}} \left( 1 - \frac{N^I N_I}{N^2} \right),
\]
where the index of \( N^I \) is lowered (raised) by \( h_{IJ} \) \( (h^{IJ}) \). With the aid of the above equation, the time slicing condition now can be expressed as
\[
B(t) = a(t)^{-6} \iff 3A + \nabla^2 \chi = \log \left( 1 - \frac{N^I N_I}{N^2} \right).
\]

The number of dynamical d.o.f. in the UG can be counted as follows: since \( N \) and \( N^I \) are just auxiliary fields and can be determined in terms of \( h_{ij} \) after algebraically solving the constraint equations \( \delta S/\delta N = 0, \delta S/\delta N^i = 0 \), (2.103) implies that the two scalar functions \( A \) and \( \chi \) are not independent. So henceforth we shall always regard \( \chi \) as being expressed (perturbatively) in terms of the other perturbations in the metric and select \( A, C_i \) and \( D_{ij} \) as the dynamical fields, the number of which is in agreement with that in the SFSG.

We are now in a position to expand the effective action (2.96) up to the quadratic level. To do this, we need to solve in the UG the constraint equations \( \delta S/\delta N = 0, \delta S/\delta N^i = 0 \) and the time slicing condition (2.103) up to the linear order in fields and plug back into (2.96). We find that
\[
S^{(2)}_T = \int dt \int \frac{M_{Pl}^2}{8} a(t)^3 \left( \ddot{b}_{ij}^2 - \frac{k^2}{a^2} D_{ij}^2 \right),
\]
\[
S^{(2)}_V = \int dt \int \frac{a(t)^3}{\dot{a}} \left( -\frac{M_{Pl}^2 \dot{H} a(t)^2}{1 - 4H a^2/k^2} C_i^2 \right),
\]
\[
S^{(2)}_S = \int dt \int \frac{a(t)^3}{\dot{a}} \left[ \frac{9M_{Pl}^3 \dot{H} a(t)^2}{4(3H a^2 - k^2)} \left( \dot{A} - \frac{\dot{H}}{H} A \right)^2 - \frac{9M_{Pl}^3 \dot{H}}{4} \left( 1 + \frac{\ddot{H}}{3H \dot{H}} \right) A^2 \right].
\]
where the symbol \( \int \frac{dk}{(2\pi)^3} \) is abbreviation for \( \int \frac{dk}{(2\pi)^3} \) and where we have used the Friedmann equation
\[
\ddot{\rho} = -F(\dot{B}) = 3M_{Pl}^2 H^2, \quad \dot{\rho} + \dot{\rho} = -2F'(\dot{B}) \dot{B} = -2M_{Pl}^2 \dot{H}.
\]

\(^{12}\)Recall that both \( \vec{x} \) and \( \vec{\phi} \) are vectors in the diagonal \( SO(3) \) group, so we need not to distinguish the spatial index “i, j, ...” from the internal ones “I, J, ...”. 

---

\( h^{ik} h_{kj} = \delta^i_j \).
CHAPTER 2. EFT CONSTRUCTION FOR FLUIDS

Notice that in the long wavelength limit $\sigma \equiv k/aH \rightarrow 0$, the actions (2.104—2.106) reduce to

$$S^{(2)} = S_{T}^{(2)} + S_{V}^{(2)} + S_{S}^{(2)} \simeq \int dt \int k M_{P}^{2} a(t)^{3} \left[ \frac{1}{8} D_{ij}^{2} + \frac{1}{4} \left( k C_{i} \right)^{2} + \frac{3}{4} A^{2} \right] \quad (2.108)$$

A few comments are in order: i) The speed of a sound wave mode in curved spacetime is given by

$$c_{s}^{2} = \left. \frac{dp}{d\rho} \right|_{B=B} = -1 - \frac{\ddot{H}}{3H\dot{H}}$$

and that of a transverse mode by $c_{T}^{2} = 0$. For $c_{s}$ to be sub-luminal (and real), we need $1 < -\ddot{H}/3H\dot{H} < 2$. As a direct consequence, a perfect fluid (2.96) can not drive inflation. To see this, let us for simplicity consider an equation of state with constant $w = p/\rho$, the inequality above implies that the scale factor evolves as $a(t) \propto t^{n}$, with $\frac{1}{3} < n < \frac{2}{3}$, whereas we need $n > 1$ or $a(t) \propto \exp (Ht)$ for the universe to inflate. Indeed to have inflation, we need to relax some symmetry requirements for our system, using less symmetric objects (e.g. a solid) as the matter content, which will be discussed in Chapter 4. Another reason why we cannot use (2.96) for inflation model is its problematic quantum mechanical feature associated with the vortices, as we discussed in Section 2.3.

ii) From the IR quadratic action (2.108), we know that it is appropriate to treat $A \sim k C_{i} \sim D_{ij}$ as of the same order in the spatial gradient expansion. Also, despite the deceptive appearance, $\nabla^{2}\chi$ is not necessarily of higher order than $A$ in the spatial gradient expansion; in fact as we will show soon, they are both of the leading order. Thus the scalar, vector and tensor metric perturbation in (2.98) are actually all of the same order in the spatial gradient expansion, i.e.

$$A \sim k^{2} \chi \sim k C_{i} \sim D_{ij} \sim O \left( \sigma^{0} \right) , \quad \sigma \equiv \frac{k}{aH} \quad (2.110)$$

2.4.2 The Conserved Curvature Perturbation to Non-linear Order

As mentioned in the previous subsection, if the matter content of the universe can be modeled as an ordinary perfect fluid with no conserved charge, there is only one dynamical scalar field. In the UG it is parametrized by $A(x)$, the (dynamical) scalar perturbation in the spatial metric (c.f. Eqn. (2.98)). In this subsection, we will show that the evolution of $A$
will remain time-independent as long as the mode is outside the horizon. The conservation is preserved up to all orders in the field expansion; at the linear order, \(A\) coincides with (twice) the usual (linear) curvature perturbation. Therefore we shall define our nonlinear curvature perturbation \(\zeta\) as

\[
\zeta = A/2. \quad (2.111)
\]

The proof of the conservation of \(\zeta\) on large scales will proceed in two main steps as follows: Firstly we expand the action (2.96) up to the leading order in the spatial gradient, — i.e. up to the order \(O(\sigma^0)\) (2.110), — while keeping all orders in fields. We show that the IR action starts with terms involving at least two time derivatives, each of which acts on a different field, and hence the constant (time-independent) configurations of \(A, C_i, D_{ij}\) are allowed solutions of the classical equations of motion. And secondly we show that these constant configurations are stable under small fluctuations — i.e. they are indeed attractors. As we will see, the proofs for the scalar, vector and tensor perturbations are essentially identical.

**Step 1:** To show the conservation of \(A, C_i, D_{ij}\), we expand the Lagrangian (2.96) up to the first order in the temporal derivative and to the zeroth order in the spatial gradients, while keeping all orders in fields. The main challenge of doing this is to express the non-dynamical quantities — such as \(N, N^i\) and \(\chi\) — in terms of the dynamical ones, by solving the time slicing condition (2.103) and the constraint equations which are given by

\[
0 = \frac{\delta S}{\delta N} = \frac{M_{Pl}^2}{2} \left[ R^{(3)} + N^{-2} \left( 6H^2 + 2H \text{Tr} \dot{M} + \frac{1}{4} \text{Tr} \left( \frac{d}{dt} e^{-M} \frac{d}{dt} e^M \right) + \frac{1}{4} \left( \text{Tr} \dot{M} \right)^2 \right) + \nabla_i N^j (\ldots) \right] + F(B) + 2F'(B)B \frac{N^i N_i}{N^2 - N^j N_j}, \quad (2.112)
\]

\[
0 = \frac{\delta S}{\delta N^i} = \nabla_i \left[ \frac{-2H}{N} - \frac{\text{Tr} \dot{M}}{2N} + \frac{\nabla_k N^k}{N} \right] + \nabla_j \left[ \frac{1}{2N} \left( e^{-M} \frac{d}{dt} e^M \right)_i^j - \frac{\nabla^j N_i + \nabla_i N^j}{2N} \right] - 2F'(B)B \frac{N N_i}{N^2 - N_k N^k}, \quad (2.113)
\]

where \(R^{(3)}\) is the spatial Ricci scalar constructed from \(h_{ij}\), \(\nabla\) the covariant derivative compatible with \(h_{ij}\), and the matrix \(M_{ij} = \log(h_{ij}a^{-2})\). The “…” in the first equation stands for terms that are regular in the limit \(\sigma \to 0\) — the form of which is irrelevant in the following analysis. Inspecting the constraint equation (2.113), one finds that \(N^i\) starts at
least at the order $O(\sigma)$. Thus, as long as the super-horizon modes are concerned, all the terms in (2.112) containing $N^i$ or $\nabla$ (as well as $R^{(3)}$) can be neglected, which leads to a crucial fact that $\delta N \equiv N - 1$ in the super-horizon regime starts with terms involving two time derivatives:

$$\delta N = \left[ 1 + \frac{1}{24H^2} \text{Tr} \left( \frac{d}{dt} e^{-M} \frac{d}{dt} e^M \right) \right]^{1/2} - 1, \quad \sigma \to 0 \quad (2.114)$$

Moreover, we see that upon each field, there is at most one time derivative acting. On the other hand, the time slicing condition (2.103), in this long wavelength limit, reduces to

$$\text{Tr} M = 3A + \nabla^2 \chi = O(\sigma^2), \quad (2.115)$$

which verifies our claim that $\nabla^2 \chi$ is of the same order as $A$ in the spatial gradient expansion.

It then follows immediately that in the super-horizon regime, the Lagrangian (2.96) becomes

$$\lim_{k/aH \to 0} \mathcal{L} \simeq 2a(t)^3 N F(B) \simeq -6a(t)^3 M_{Pl}^2 H^2 + \sum_{n \geq 2} \mathcal{L}_n(\phi_a) \quad (2.116)$$

In the last step we have denoted the dynamical fields — $A,C_i,D_{ij}$ — collectively by $\phi_a$ and the number of fields contained in $\mathcal{L}_n$ by the subscript “$n$”. Notice that schematically $\mathcal{L}_n$’s take the form of

$$\mathcal{L}_2 \sim Q_2(t) G_{2b}^{ab} \dot{\phi}_a \dot{\phi}_b, \quad \mathcal{L}_{n \geq 2} \sim Q_n(t) G_{nkl}^{ab} \cdots (\phi) \dot{\phi}_a \cdots \dot{\phi}_k \dot{\phi}_l \quad (2.117)$$

with $Q_n(t)$ being a function of time consisting of $a(t), H$ etc.. Besides the irrelevant field-independent term, this long-wavelength Lagrangian starts at two time derivative level and hence the e.o.m. following from it reads

$$f_1(\phi, \dot{\phi}) \ddot{\phi}_a + f_2(\phi, \dot{\phi})_a \dot{\phi}_b = 0 \quad (2.118)$$

Therefore it indeed admits $A, C_i, D_{ij} = \text{constant}$ as solutions of the classical equation of motion. Since this long wavelength Lagrangian contains all orders in fields, the conservation of these fields on large scales must be preserved nonlinearly.

**Step 2:** Now we show the solutions $A, C_i, D_{ij} = \text{constant}$ are actually attractors. We just work on the scalar case, since the analysis for the other two is essentially identical. Writing
A as \( A = A_0 + \delta A \), owing to the constancy of \( A_0 \), the quadratic action for the fluctuation \( \delta A \) in the long-wavelength limit takes the same form as that for \( A \), which is given by

\[
\delta S_s^{(2)} = \int dt \int \mathbf{k} \frac{3M_{Pl}^2}{4} a(t)^3 \delta A^2
\]

from which follows the linearized equation of motion for \( \delta A \)

\[
\ddot{\delta A} + 3H \dot{\delta A} = 0 \tag{2.119}
\]

It admits two general solutions — one decaying mode and one constant mode:

\[
\delta A_1 = \int \frac{dt}{a(t)^3}, \quad \delta A_2 = \text{const.} \tag{2.120}
\]

This thus confirms that the solution \( A = \text{constant} \) is an attractor.

As we saw, our approach is more powerful in some respect than those in previous literature, for it enables us to show that the vector perturbation \( C_i \) and the tensor perturbation \( D_{ij} \) are also conserved on super-horizon scales in the same manner as their scalar counterpart \( A \). In next subsection, we will show that the nonlinear, conserved \( \zeta \equiv A/2 \) and \( D_{ij} \) we constructed here agrees with that in [10, 16], up to a time-independent spatial coordinates redefinition.

### 2.4.3 Remarks on the Conservation of \( \zeta \)

As was illustrated in the last subsection, we have constructed the curvature perturbation \( \zeta \) (to nonlinear orders) through two steps: \( i \) choose a coordinate system such that the spatial coordinates comove with the fluid and that equal time slices coincide with the uniform density slices; \( ii \) define \( \zeta \) to be half the coefficient of the term in \( \log \left( a^{-2} g_{ij} \right) \) that is proportional to \( \delta_{ij} \) — that is,

\[
\zeta = \frac{1}{4} \nabla^{-2} \left( \delta_{ij} \nabla^2 - \partial_i \partial_j \right) \log \left[ a^{-2} g_{ij} \right] \tag{2.121}
\]

We showed that the \( \zeta \) thus defined is time-independent in the super-horizon regime to all orders in fields.

Our result of the conservation of the non-linear curvature perturbation \( \zeta \) on large scales agrees with previous literature [10,16]. But in our analysis we did not neglect the presence
of the vector and tensor perturbations, nor did we use the property of local homogeneity and isotropy. As we showed explicitly, the spatial metric in UG on large scales ($\sigma \to 0$) is given by

$$g_{ij} = a(t)^2 \exp \left[ 6\zeta \left( \frac{1}{3} \delta_{ij} - \hat{k}_i \hat{k}_j \right) + \text{vector + tensor} \right].$$

(2.122)

At a glimpse, one may think that even though the vector and tensor perturbations can be neglected, this metric is anisotropic, which is inconsistent with our intuition about cosmological fluids that no anisotropy in the metric will be generated if the fluid is free of anisotropic stress. However, this puzzle results from our non-conventional choice of coordinates: we can perform a further coordinate transformation (from the UG)

$$t \to t, \quad x^I \to y^I = x^I + \xi^I(\vec{x})$$

(2.123)

such that

$$\phi'^I(t, \vec{y}) = y^I + \xi'^I(\vec{y}), \quad \text{and } g'_{ij}(t, \vec{y}) = a(t)^2 e^{2\zeta'(\vec{y})} \exp \gamma'_{ij}(\vec{y}), \quad \text{with } \partial_i \gamma'_{ij} = \gamma'_{ii} = 0$$

(2.124)

where the time-independence of $\xi'(\vec{x})$ follows directly from the time-independence of $A, C_i,$ and $D_{ij},$ and where the function $\zeta'(\vec{y})$ is obtained by inverting the 3-diffeomorphism $y^I(\vec{x}) = x^I + \xi^I(\vec{x})$. Since $\zeta$ and $\zeta'$ are related by a (time-independent) coordinate transformation (2.123), the super-horizon conservation for one implies that for the other.

Notice that a coordinate transformation like (2.123) alters neither the time slices nor the threading of the spatial coordinates, the latter of which because

$$u'^I \propto e^{\alpha\beta\gamma} \partial_\alpha \phi'^I \partial_\beta \phi'^2 \partial_\gamma \phi'^3 = 0, \quad u'_I = g_{0I} u^0 = h_{ij} N^j u^0 \sim O(\sigma).$$

(2.125)

Therefore this implies that in the new coordinate system and in the super-horizon regime, the spatial components of the energy momentum tensor $T_{ij}$ (c.f. Eqn. (2.7))— which is reduced to $T_{ij} = pg_{ij}$ on large scales — is isotropic. Were the tensor perturbation ignored,

13In fact having fixed the UG by $\phi' = x^I$ and $B = a(t)^{-6},$ we have depleted all gauge freedoms for choosing a coordinate system. Furthermore, as we argued earlier, we should not think that the terms in $g_{ij}$ with “superficially” more derivatives such as $\partial_i \partial_j \chi,$ $\partial_i C_j$ are of higher order in the spatial gradient expansion than $A \delta_{ij}.$ As a consequence of this, once the UG is chosen, we are not entitled to assume any property about $g_{ij}$ on large scales.
the spatial metric $g_{ij}$ would also become homogeneous and isotropic, the former of which following from the fact that in the limit $\sigma \to 0$ the spatial dependence of $\zeta'$ is negligible.

In summary, although convenient for computational purposes, the UG coordinate is not capable of exhibiting the local homogeneity and isotropy in the long wavelength limit. Fortunately there exists a new coordinate system, which is indistinguishable from the UG coordinates by only inspecting the physical quantities such as $\rho, p, u^i$ etc. (The invariance of $u^i$ or $u_i$ under the coordinate transformation (2.123) was shown explicitly in (2.125). While the invariance of $\rho$ and $p$ follows directly from the fact that, in the UG, $\rho$ and $p$ are only functions of time, which is untouched in coordinate transformation (2.123)). And in such new coordinates the existence of local homogeneous and isotropic patches on large scales can be verified (In contrast, many previous papers adopted this as one of the key assumptions.). The $\zeta'$ in (2.124) is also time-independent on large scales and it agrees perfectly with the definition in [10,16].

Before ending this subsection, we want to point out a subtlety about the vector perturbation (associated with the vortex degree of freedom of the fluid). As we showed in last section, on super-horizon scales, the vector field $C_i$ is time-independent. However, this field does not correspond to a physical vector mode, as we will now explain. Let’s transform to the coordinate system specified by (2.124). Owing to the internal volume-preserving-diffeomorphism invariance (2.4) of the fluid system, the configuration of Eqn. (2.124) is in fact physically equivalent to the one free of any vortex degrees of freedom (divergence-free vector modes), either in the matter fields or in the dynamical metric components. That is, as the leading order contribution in the spatial gradient expansion, the constant (in time) $C_i$ configuration has no physical significance; the well known decaying vector modes in the universe dominated by perfect fluids come from subdominant, time-dependent terms in $g_{ij}$ in the UG.
Chapter 3

EFT Approach to Gravity Theories: A Modification of Einstein Gravity

3.1 An Effective Field Theoretic Point of View of General Relativity

At classical level, General Relativity is widely believed to be the correct gravity theory, since its predictions are in agreement with all the local gravity tests. The equations of motion for the dynamical metric—the Einstein equations—are obtained by varying with respect to the metric the Einstein-Hilbert action

\[ S_{EH} = \frac{M_{Pl}^2}{2} \int d^4x \sqrt{g} \, R, \]  

(3.1)

where \( R \) is the Ricci scalar constructed from \( g_{\mu\nu} \) and its inverse \( g^{\mu\nu} \).

As a quantum theory, however, it is not renormalizable. To see this, let us expand the Einstein Hilbert action \((3.1)\) around the flat Minkowskian background:

\[ g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \quad \text{with } \eta = \text{diag} \{-1, 1, 1, 1\}, \]  

(3.2)

where \( h_{\mu\nu} \) transforms as a massless spin-two field under the isometries of the background
metric — the Poincaré group. Up to the quadratic level, we get

\[ L_2 = \frac{M_{Pl}^2}{4} \left( -\partial_\mu h^{\mu\nu} \partial_\nu h + \partial_\mu h^{\mu\alpha} \partial_\nu h^{\nu}_\alpha - \frac{1}{2} \partial_\mu h^{\alpha\beta} \partial^\mu h_{\alpha\beta} + \frac{1}{2} \partial_\mu h \partial^\mu h \right) \]  

(3.3)

where the indices of \( h_{\mu\nu} \) and \( \partial_\mu \) are raised with the flat background (inverse) metric \( \eta^{\mu\nu} \). Also we have dropped total derivatives and defined \( h \equiv \eta^{\mu\nu} h_{\mu\nu} \).

If we go further to derive the interaction terms, schematically they take the form

\[ L_n \sim \frac{1}{M_{Pl}^{n-2}} (\partial h)^2 h^{n-2} , \quad n \geq 3. \]  

(3.4)

where we have canonically normalized \( h_{\mu\nu} \) to be \( h_{\mu\nu} = 2M_{Pl}^{-1} \hat{h}_{\mu\nu} \). At the external momentum/energy \( p \sim M_{Pl} \), strong self-interactions in the Einstein-Hilbert actions are aroused, signifying that this theory breaks down in the UV. This can be seen by checking the \( 2 \rightarrow 2 \) graviton scattering: at external momenta \( p \), the amplitude goes like \( p^2/M_{Pl}^2 \). Therefore when \( p \sim M_{Pl} \), the amplitude is of the order unity and unitarity is violated.

So we should really regard the General Relativity as an effective field theory of a massless spin-2 field, with a cut-off scale \( \sim M_{Pl} \simeq 10^{18} \text{GeV} \). At the energy level of experimental/observational interest, it behaves like a good quantum theory. And like other EFTs, we can use it to study low energy phenomena without being affected by the existence of UV degrees of freedom. For instance, in \cite{[44],[45]}, it was showed that using the Einstein-Hilbert action as an effective field theory, one can extract information about the quantum corrections to the gravitational potential between point masses, which only involves the low energy portion of the theory.

Perhaps more strikingly, we can apply this logic reversely — which is also more in parallel with the spirit of the EFT construction discussed in previous sections — by asking what is the most general low energy theory for a massless spin-2 field. It turns out that the answer is that, to lowest order in derivatives, such a theory is uniquely General Relativity, whose action is precisely given by (3.1).

In the following, we will present some simple argument for this. The full discussion can be found in Ref. \cite{[46]}. First, notice that the quantum theory for massless fields enjoys the so-called gauge invariance, which, roughly speaking, is an equivalence condition imposed on different field configurations, so that the number of dynamical quantum fields is the same
as that of the polarization of the quantum states \[47\]. For instance, for a free massless
spin-1 fields \( A_\mu(x) \), we should impose a gauge redundancy such that the following two field
configurations are considered identical:

\[
A_\mu(x) \sim A'_\mu(x) = A_\mu(x) - \partial_\mu \lambda(x) , \quad \text{with } \lambda(x) \text{ an arbitrary function.}
\]

while for a free massless spin-2 field \( h_{\mu\nu} \), the gauge redundancy condition (in flat Minkowskian
background) is

\[
h_{\mu\nu} \sim h'_{\mu\nu} = h_{\mu\nu} - \partial_\mu \xi_\nu - \partial_\nu \xi_\mu .
\]

The free theory action for \( A_\mu(x) \) or \( h_{\mu\nu}(x) \) should be understood as a functional of such
equivalence classes — the class of field configurations related by gauge transformations —
rather than of a single field configuration. In more familiar field theoretic language, these
quadratic actions are required to be invariant under the gauge transformations specified by

\[
A_\mu \rightarrow A'_\mu = A_\mu - \partial_\mu \lambda , \quad (3.7)
\]

\[
h_{\mu\nu} \rightarrow h'_{\mu\nu} = h_{\mu\nu} - \partial_\mu \xi_\nu - \partial_\nu \xi_\mu . \quad (3.8)
\]

for spin-1 and spin-2 field respectively. This requirement is enough to determine the form
of the actions up to an overall normalization; the free theory action for a spin-1 field \( A_\mu \) is
just the free Maxwell theory

\[
S_2[A] = -\frac{1}{4} \int d^4 x F_{\mu\nu} F^{\mu\nu} , \quad \text{with } F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu .
\]

and that for a spin-2 field \( h_{\mu\nu} \) is given by \((3.3)\).

Now we are interested in adding nonlinear self-interaction to the free graviton action
\((3.3)\). To consistently achieve this, we claim that we should promote the free gauge trans-
formation \((3.8)\) to the full gauge redundancy transformation:

\[
h_{\mu\nu} \rightarrow h'_{\mu\nu} = h_{\mu\nu} - \partial_\mu \xi_\nu - \partial_\nu \xi_\mu - \xi^\rho \partial_\rho h_{\mu\nu} - \partial_\mu \xi_\rho h_{\rho\nu} - \partial_\nu \xi_\rho h_{\rho\mu} + O(\xi^2) .
\]

and at lowest order in derivatives, the perturbative interacting theory for \( h_{\mu\nu} \) invariant
under \((3.10)\) can be resummed to give the Einstein-Hilbert action \((3.1)\). We will illustrate
this idea by examining some analogous cases.
CHAPTER 3. EFT APPROACH TO GRAVITY THEORIES: A MODIFICATION OF EINSTEIN GRAVITY

From Maxwell Theory to QED

In this section we consider how to couple a single spin-1 field to a matter sector, which is taken as the free Femionic theory

$$S[\Psi] = \int \! d^4 x \left( i \bar{\Psi} \gamma^\mu \partial_\mu \Psi - m \bar{\Psi} \Psi \right). \quad (3.11)$$

This possesses a global $U(1)$ symmetry

$$\Psi \rightarrow e^{-i\alpha} \Psi, \quad \bar{\Psi} \rightarrow \bar{\Psi} e^{i\alpha}, \quad \alpha \ \text{const.}, \quad (3.12)$$

with the associated Noether current given by

$$j^\mu_\Psi = e \bar{\Psi} \gamma^\mu \Psi. \quad (3.13)$$

Now turn to construct an interaction term $L_{\text{int}}$ between the gauge field and the fermion fields that is also invariant under the gauge transformation (3.7). A natural candidate is

$$S_{\text{int}} = \int \! d^4 x \ A_\mu j^\mu_\Psi. \quad (3.14)$$

then under the gauge transformation, the variation

$$\delta S_{\text{int}} = + \int \! d^4 x \ \lambda(x) \partial_\mu j^\mu_\Psi \quad (3.15)$$

vanishes on shell. That is, the action $S[A] + S_{\text{int}}[A, \Psi] + S[\Psi]$ has a on-shell gauge symmetry: $A_\mu \rightarrow A_\mu - \partial_\mu \lambda$. Notice that till now we didn’t involve the influence of the gauge redundancy on the matter sector; if we do, a miracle occurs. Promote the global $U(1)$ transformation parameter $\alpha$ in (3.12) to be a spacetime-dependent function: $\alpha \rightarrow \alpha(x)$ and consider the transformation

$$\Psi \rightarrow e^{-i\alpha(x)} \Psi, \quad \bar{\Psi} \rightarrow \bar{\Psi} e^{i\alpha(x)}, \quad A_\mu \rightarrow A_\mu - \partial_\mu \lambda(x). \quad (3.16)$$

The variation of $S$ is given by

$$\delta S = + \int \! d^4 x \ \lambda(x) \partial_\mu j^\mu_\Psi - \int \! d^4 x \ \alpha(x) \partial_\mu j^\mu_\Psi, \quad (3.17)$$

where the second term in the r.h.s. comes from the variation of $S[\Psi]$, following the standard Noether procedure. We have also used the fact that $j^\mu_\Psi$ is unaltered under (3.16). Obviously $S$ is invariant off-shell if we identify $\alpha(x) = \lambda(x)$. We then recognize that Eqn. (3.16) becomes the familiar $U(1)$ gauge transformation and $S$ is just the interacting theory between the gauge field $A$ and the fermion field $\Psi$. 
Pure Yang-Mills Theory

In this subsection, we want to apply the logic to Yang-Mills theory, i.e. to construct Yang-Mills type interaction among gauge fields via the gauge redundancy condition.

Let start with \( N \) copies of free \( U(1) \) gauge field action

\[
L_2 = \sum_{a=1}^{N} \left[ -\frac{1}{2} \partial_\mu A_\mu^a \partial_\nu A_\nu^a + \frac{1}{2} \partial_\mu A_\nu^a \partial_\nu A_\mu^a \right],
\]

(3.18)

where each \( A_\mu^a \) is a spin-1 field. This free action is invariant under free gauge transformation

\[
A_\mu^a \to A_\mu^a - \partial_\mu \theta^a(x), \quad a = 1, \ldots, N.
\]

(3.19)

In addition, one can check that \( L_2 \) possesses some global symmetry

\[
A_\mu^a \to A_\mu^a - g f^{abc} A_\mu^b \alpha^c, \quad \alpha^c \text{ const.},
\]

(3.20)

where \( f^{abc} \) is totally anti-symmetric under swapping any pair of indices and it satisfies the Jacobi Identity

\[
- f^{abd} f^{cde} + f^{cbd} f^{ade} = f^{acd} f^{dbe}.
\]

(3.21)

The Noether current of this global symmetry is given by

\[
j^{a\mu} = g f^{abc} A_\nu^b (-\partial_\mu A_\nu^c + \partial_\nu A_\mu^c).
\]

(3.22)

That is, the \( f^{abc} \)'s furnish the adjoint representation of some Lie group, which is an input symmetry at the level of free field theory (as the global global \( U(1) \) symmetry in the QED case).

Now we are in a position to add consistently interaction terms among the \( A_\mu^a \)'s. Notice that in this case, there is no matter sector involved. However, for each fixed \( a \), say \( a = 1 \), we can regard the other gauge fields \( A_\mu^b, b = 2, \ldots, N \) as the analogous ‘matter sector’ for \( A_\mu^1 \). So following the same logic as in last subsection, we propose the interaction term for \( A_\mu^1 \) to be

\[
L_{\text{int}} \supset A_\mu^1 j^{1\mu}, \quad \text{with } j^{1\mu} = \sum_{\hat{b}, \hat{c}=2}^{N} g f^{1\hat{b}\hat{c}} A_\nu^{\hat{b}} (-\partial_\mu A_\nu^{\hat{c}} + \partial_\nu A_\mu^{\hat{c}}).
\]

(3.23)

Notice that \( j^{1\mu} \) does not contain the \( A_\mu^1 \) field, so under the free gauge transformation

\[
A_\mu^1 \to A_\mu^1 - \partial_\mu \theta^1(x), \quad j^{1\mu} \text{ remains unchanged while the interaction term, } A_\mu^1 j^{1\mu}, \text{ is invariant only on shell.} 
\]
For this to be lifted to an off-shell symmetry transformation, it is necessary for the $A_1^1$ gauge transformation and the global transformation of $A_{\tilde{a}}^a$ fields (with $\tilde{a} = 2, ..., N$) to conspire with each other to give the adjoint gauge transformation for $A_1^1$:

$$A_1^1 \rightarrow A_1^1 - \partial_\mu \theta^1(x) , \quad A_{\tilde{a}}^a \rightarrow A_{\tilde{a}}^a - \sum_{b=2}^{N} gf^{\tilde{a} b 1} A_{\nu}^b \theta^1(x) , \quad \tilde{a} = 2, ..., N.$$ 

(3.24)

In other words, we have promoted the global symmetry parameter $\alpha^1$ in (3.20) to be spacetime-dependent and set it equal to $\theta^1(x)$ in (3.19): $\alpha^1 \rightarrow \alpha^1(x) = \theta^1(x)$. On the other hand, there is nothing special about $A_1^1$; we can repeat the above procedure for all $A_a^a$'s. Hence we recover the familiar full gauge transformation

$$A_a^a \rightarrow A_a^a - \partial_\mu \theta^a(x) - gf^{abc} A_b^b \theta^c(x) , \quad a = 1, ..., N .$$ 

(3.25)

An action invariant under this full gauge transformation (3.25) must satisfy the following recursive relation:

$$\delta S[A] = 0 \Rightarrow \partial_\mu \left( \frac{\delta S[A]}{\delta A_{a}^a} \right) = -gf^{abc} A_{b}^b \frac{\delta S[A]}{\delta A_{c}^c}$$ 

(3.26)

and all the interaction terms can be obtained by induction:

$$\partial_\mu \left( \frac{\delta S_n[A]}{\delta A_{a}^a} \right) = -gf^{abc} A_{b}^b \frac{\delta S_{n-1}[A]}{\delta A_{c}^c} , \quad \text{for } n \geq 3 ,$$ 

(3.27)

where $S_n[A]$ ($n \geq 3$) denotes $n$-th interaction term involving $n$ $A_a^a$ fields. For pure non-Abelian Yang Mills theory, the induction completes at $n = 4$ level; the resulting interacting action constructed from (3.18) can be written in a compact (and familiar) form

$$\mathcal{L} = -\frac{1}{4} F_{\mu \nu}^{a} F^{a \mu \nu} , \quad \text{with } F_{\mu \nu}^{a} = \partial_\mu A_{\nu}^a - \partial_\nu A_{\mu}^a + gf^{abc} A_{\mu}^b A_{\nu}^c .$$ 

(3.28)

From Free Spin-2 Field Theory to General Relativity

Finally we come to the construction of General Relativity out of the quadratic action (3.3). First it is easy to check that this quadratic action admits a global symmetry given by the transformation (spacetime translations)

$$h_{\mu \nu} \rightarrow h_{\mu \nu} - \theta^\rho \partial_\rho h_{\mu \nu} , \quad \theta^\rho \text{ const.}$$ 

(3.29)
The associated Noether current is just the energy momentum tensor of the quadratic action $T^2_{\mu\nu}$. As in last subsection, to construct consistently self-interaction Lagrangian of $h_{\mu\nu}$, we need to combine the free gauge transformation (3.8) and the global symmetry transformation (3.29) and identify the parameter. This leads to the usual gauge transformation for $h_{\mu\nu}$ (3.10).

Then the interaction Lagrangian invariant under the full gauge transformation can be constructed order by order via some recursive condition, as we did for the Yang-Mills theory. However, here we can employ a trick that does automatically the infinite series resummation for us. Note that till now, we only talked about the field theory for spin-2 field $h_{\mu\nu}$ in a flat Minkowskian spacetime. Defining a field $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$, the transformation (3.10) can be rewritten as

$$g_{\mu\nu}(x) \rightarrow g'_{\mu\nu}(y) = \partial x^\alpha \partial x^\beta g_{\alpha\beta}(x), \quad \text{with } y^\mu(x) = x^\mu + \xi^\mu(x). \quad (3.30)$$

If we invoke the geometric description of $g_{\mu\nu}$ and interpret it as the metric of some curved spacetime, it follows immediately that, to lowest order in derivatives, the theory of the metric field $g_{\mu\nu}$ that is invariant under the general covariance (3.30) is just the General Relativity (3.1).

### 3.2 A Modification of GR in the Infrared: Motivations and Examples

In the last section, we argued that the GR is unambiguously the low energy effective theory for a massless spin-2 field. Why do we want to modify it, given the fact that it is such a rigorous theory and is in agreement with all local gravity test?

The main motivation comes from the existence of a tiny Cosmological Constant (CC). According to supernova data, our universe is undergoing an acceleration in its expansion. To explain this within the framework of GR as the gravity theory, the simplest solution is to introduce a constant $\Lambda$ in Einstein equations (or a term $\sqrt{-g}\Lambda$ in the EH action.) This CC is often regarded as a new type of ‘matter content’ in the universe known as dark energy, which is responsible for the accelerating expansion, and its energy density is measured to be
\[ \rho_\Lambda \sim 10^{-11} \text{eV}^4. \]

It is curious to notice that this energy scale is much smaller than that for the vacuum energy we would estimate naively by cutting off zero-point energies at \( E \sim M_{Pl} \):

\[
\frac{\rho_\Lambda}{\rho_{\text{vac}}} = \frac{\rho_\Lambda}{M_{Pl}^4} \sim 10^{-120} .
\] (3.31)

This unnaturally smaller quantity is hardly a pathology of the theory, but rather is a puzzle about the lack of standard particle physics energy scales corresponding to the CC and about the miracle cancellation among the different vacuum energies. So, needed to achieve its smallness there are attempts that try to modify GR at cosmological scales (infrared) in such a way that the accelerating universe can be produced without the CC.

Since GR is the unique low energy theory for a single spin-2 field, the only possibility to modify it is to introduce new degrees of freedom or to include higher derivatives of \( g_{\mu\nu} \). In many circumstances these two possibilities are in fact equivalent. For instance, Ref. [?] showed that replacing the EH Action with \( F(R) \) — a general function of the Ricci scalar — is equivalent to adding a scalar field to the GR theory. Therefore it is quite natural to start our journey to modify gravity from the simplest case: the scalar tensor theories — the GR theory plus a single scalar degree of freedom.

Before presenting examples of scalar-tensor theories of great interests, we want to point out a caveat which may be relevant to every modification of gravity theory. One should expect that all deviations from GR occur on cosmological scales, while on short scales — the scale of the Solar system or of our Galaxy — GR should be recovered. In other words, all additional degrees of freedom need to conceal themselves from detection so that the local gravity tests are satisfied. Any attempt of modification that fails this is unqualified as an alternative gravity theory.

The mechanism for hiding the extra scalar degree of freedom is usually referred to as the ‘screening mechanism’, which differs case by case and is often regarded as the essence of each modified gravity model. The two models that we are going to review implement the two best known screening mechanisms — the ‘chameleon’ mechanism and the ‘Vainshtein’ mechanism — which result from very different theoretical considerations.
The Chameleon Model and Chameleon Mechanism

The chameleon model [23,24] involves an extra scalar field, nonlinearly interacting with itself via a non-derivatively coupled potential. The chameleon mechanism operates whenever the scalar field couples to matter in such a way that the effective mass of the scalar field depends on the mass density of its environment: deep in the space where the mass density is very low (approximately the mean density of the universe), the scalar is light and mediates a long-ranged fifth force of gravitational strength, which could potentially have an impact on observations on cosmological scales. While within some highly dense region, such as in the Solar system or in the Milky Way, it acquires a large mass. Therefore the scalar (fifth) force becomes Yukawa suppressed beyond a short range and hence is not detectable.

Explicitly, the action of the chameleon theory is given by

$$S = \int d^4x \sqrt{-g_E} \left( \frac{R_E}{16\pi G_N} - \frac{1}{2} (\partial \phi)^2 - V(\phi) \right) + S_m[A^2(\phi)g^E_{\mu\nu}, \Psi].$$  \hspace{1cm} (3.32)

where the superscript ‘E’ is to remind us that these quantities are evaluated in the Einstein frame. Matter fields described by $\Psi$ couple to $\phi$ through the Jordan frame metric $g^J_{\mu\nu}$ which is related to the Einstein frame metric $g^E_{\mu\nu}$ through some (positive) conformal factor $A^2(\phi)$:

$$g^J_{\mu\nu} = A^2(\phi) g^E_{\mu\nu}. \hspace{1cm} (3.33)$$

Although the two frames are physically equivalent, cosmological observations are implicitly performed in Jordan frame, where the masses of particles are constant. Meanwhile computation is usually simpler in the Einstein frame, since the Einstein equations take the same form as in GR. The form of the scalar potential does not play a vital role in the screening mechanism, as long as it satisfies certain runaway condition [23,24]. The constraints on the scalar factor $A(\phi)$ will be discussed in detail in Section ??.

The equation of motion for the scalar field is obtained by varying (3.32):

$$\Box \phi = V_{,\phi} + A_{,\phi} \rho,$$  \hspace{1cm} (3.34)

where the matter is assumed to be non-relativistic, and $\rho$ is related to the Einstein and Jordan frame matter densities by $\rho = \rho_E/A = A^3\rho_1$ — defined such that $\rho$ is conserved in
CHAPTER 3. EFT APPROACH TO GRAVITY THEORIES: A MODIFICATION OF EINSTEIN GRAVITY

58

the usual sense in Einstein frame

To have an idea of how the chameleon screening mechanism operates, let us focus on
the prototype model where the scalar potential is taken to be an inverse power law and the
scale factor, in the region of interests, to be approximately a linear function:

$$V(\phi) = \frac{M^{4+n}}{\phi^n} \quad (n \geq 1) \quad , \quad A(\phi) = 1 + \frac{\beta \phi}{M_{Pl}} .$$  (3.35)

where $M$ is some mass scale and $\beta$ is some dimensionless parameter, which, as we will see
soon, determines the strength of the fifth force on cosmological scales. To make this model
somehow distinguishable from GR, we require $\beta \sim O(1)$, so that the scalar force can be of
gravitational strength.

Let us consider a spherically symmetric overdense object sitting in some homogeneous
medium with mass density $\rho_{\infty}$, i.e. the density profile is given by

$$\rho(r) = \begin{cases} 
\rho_c & \text{for } r < R_c , \\
\rho_{\infty} & \text{for } r > R_c .
\end{cases}$$  (3.36)

Near the object, the gravitational field is weak and the e.o.m. for the scalar field (3.34) is
reduced to

$$\frac{d^2 \phi}{dr^2} + \frac{2}{r} \frac{d \phi}{dr} = \frac{dV}{d\phi} + \frac{\beta \rho(r)}{M_{Pl}}$$  (3.37)

subject to the following two boundary conditions

$$\frac{d\phi}{dr} = 0 \quad \text{at} \quad r = 0 ;$$

$$\phi \rightarrow \phi_{\infty} \quad \text{at} \quad r \rightarrow \infty .$$  (3.38)

The non-linear feature of the potential $V(\phi)$ makes it difficult to solve this differential
equation analytically over the whole domain of $r$. Fortunately, when focusing only on the
scalar profile inside or outside the object, we can apply certain approximations and extract
analytic solutions. And the full solution to (3.37) and (3.38) can be obtained by ‘smoothly
gluing’ together the interior and exterior scalar profile.

We are mostly interested in the comparison of the fifth force to the gravitational force
exerting on some test particle. It is sufficient to check only the case of a test particle moving

\footnote{More precisely: $U \cdot \nabla \rho = -\rho \nabla \cdot U$ for a pressureless fluid. Here, $\rho = \rho_c / A$, $U^\mu$ is the fluid 4-velocity in
Einstein frame ($= A \times$ Jordan frame velocity), and indices are contracted with the Einstein frame metric.}
outside but near the spherical object. The fifth force at $r \geq R_c$ is given by (omitting order unity prefactors)

$$F_\phi \sim \begin{cases} 
\frac{\Delta \phi}{\beta M_P \Phi_N} \frac{\beta G_N M_c}{r^2}, & \text{for } \Delta \phi \lesssim \beta M_P \Phi_N, \\
\frac{\beta^2 G_N M_c}{\Phi^2}, & \text{for } \Delta \phi \gtrsim \beta M_P \Phi_N, 
\end{cases}$$

(3.39)

where $M_c$ is the total mass of the object, $G_N$ the Newtonian potential of the object at the surface, and $\Delta \phi \equiv \phi(\infty) - \phi(r = 0)$ the difference of the scalar field value at infinity and at the center. Eqn. (3.39) tells us that, for an object with fixed radius, the denser it is, the more suppressed is the scalar force (relative to gravitational force) it mediates. If we invoke the geometric interpretation of the quantity $\frac{\Delta \phi}{\beta M_P \Phi_N}$ and rewrite it as

$$\frac{\Delta R}{R_c} \simeq \frac{\Delta \phi}{\beta M_P \Phi_N},$$

(3.40)

the first equation in (3.39) implies that only the mass from a thin shell (with thickness $\Delta R$) of this spherical object contributes the fifth force while the rest gets screened.

This is the chameleon screening mechanism: if a test particle is within or close to some very dense astronomical object (e.g. the Solar system), it will not feel the fifth force; on the other hand, when the particle is in an environment where the mass density fluctuates mildly around the cosmic mean density, the fifth force the particle experiences could be of the order of the gravitational force.

The Galileon Model and Vainshtein Mechanism

The Galileon model\cite{26} is a local modification of GR due to a derivatively self-coupled scalar field. It is defined on a local FRW patch where the coordinate system is chosen such that the metric is quasi-Minkowskian. The action takes the form

$$S = \int d^4x \left( \mathcal{L}_2(h) + \frac{1}{2} h_{\mu\nu} T^{\mu\nu} + \pi T^\mu_\mu + \mathcal{L}_\pi \right) + \mathcal{O}(h^3)$$

(3.41)

where $\mathcal{L}_2(h)$ is the quadratic Lagrangian for $h_{\mu\nu}$ given by (3.3), $T^{\mu\nu}$ the energy momentum tensor of the matter sector, and $\mathcal{L}_\pi$ encodes the dynamics of the $\pi$ sector. Although we are only working at the quadratic order in $h_{\mu\nu}$, $\mathcal{L}_\pi$ is supposed to include all the nonlinear self-interactions of $\pi$, which, as we will argue soon, provide the Vainshtein screening mechanism.
The Lagrangian $L_\pi$ and the equation of motion for $\pi$ are demanded to be invariant under the galileon symmetry:

$$\pi(x) \to \pi + b_\mu x^\mu + c,$$

with $b_\mu, c$ constant.

Moreover, to avoid ghosts, we want the e.o.m. for $\pi$ to contain at most the second derivatives of $\pi$. These requirements altogether are strong enough to uniquely determine the full structure of $L_\pi$. For instance, in $d = 4$ spacetime dimension, only 5 terms (plus a constant term) are allowed. They are

$$L_\pi = M_{Pl}^2 \sum_{n=0}^{5} c_n \mathcal{E}_n,$$

where we have denoted by $\Pi$ the matrix of second derivatives of $\pi$, $\Pi_{\mu\nu} \equiv \partial_\mu \partial_\nu \pi$, by $[\ldots]$ the operation of taking the trace (with respect to the flat metric $\eta$) and by $\cdot$ the standard Lorentz invariant contraction of indices. The equation of motion for the $\pi$ field can be obtained by the usual procedure. It is given by

$$E = \frac{\delta L}{\delta \pi} = \sum_{n=1}^{5} c_n \mathcal{E}_n = -\frac{T^\mu_\mu}{M_{Pl}^2},$$

where

$$\mathcal{E}_1 = 1,$n

$$\mathcal{E}_2 = (\Box \pi),$$

$$\mathcal{E}_3 = (\Box \pi)^2 - (\partial_\mu \partial_\nu \pi)^2,$n

$$\mathcal{E}_4 = (\Box \pi)^3 - 3(\Box \pi)(\partial_\mu \partial_\nu \pi)^2 + 2(\partial_\mu \partial_\nu \pi)^3,$n

$$\mathcal{E}_5 = (\Box \pi)^4 - 6(\Box \pi)^2(\partial_\mu \partial_\nu \pi)^2 + 8(\Box \pi)(\partial_\mu \partial_\nu \pi)^3 + [(\partial_\mu \partial_\nu \pi)^2]^2 - 6(\partial_\mu \partial_\nu \pi)^4.$$
where \((\partial_\mu \partial_\nu \pi)^n\) stands for the cyclic contraction.

The attractive feature about the Galileon model is that it possesses both self-accelerating solutions and the Vainshtein screening mechanism. Let us first consider self-acceleration. By a self-accelerating solution, we mean a nontrivial \(\pi\) configuration in the absence of matter that leads to a deSitter geometry in the Jordan frame (while it is a Minkowskian geometry in the Einstein frame). This relevant \(\pi\) configuration takes the form

\[
\pi_{\text{dS}} = -\frac{1}{4} H^2 x_\mu x^\mu, \quad \partial_\mu \partial_\nu \pi_{\text{dS}} = -\frac{1}{2} H^2 \eta_{\mu\nu}, \tag{3.45}
\]

where the constant \(H\) is the inverse radius of the deSitter space. Plugging the above equation into (3.44) and setting \(T^\mu_\mu = 0\), it is straightforward to show that there exists some region in the parameter space — i.e. the space spanned by \(\{c_0 \ldots c_5\}\) — in which this configuration (3.45) is guaranteed to be a solution to the e.o.m.

Now let us turn to a spherical solution around some point mass on top of the self-accelerating solution discussed above. Let \(\pi(x) = \pi_{\text{dS}} + \varphi(r)\), one can show that \(\varphi\) satisfies

\[
d_2 (\varphi'/r) + 2d_3 (\varphi'/r)^2 + 2d_4 (\varphi'/r)^3 = \frac{M}{4\pi r^3} \tag{3.46}
\]

where we have assumed that the source, with total mass \(M\), is placed at the origin, \(\rho = M\delta^3(\vec{r})\). The coefficients \(d_2, d_3, d_4\) are constructed from \(c_i\)'s and \(H\), and restricted within some region to assure the existence of such a spherical solution; in particular \(d_2, d_4\) are required to be non-negative \[26\]. Notice that Eqn. (3.46) is an algebraic equation for \(\varphi'/r\), and at very small \(r\), the cubic term on the l.h.s. dominates, while at far away from the source \(r \to \infty\), the linear terms dominates. It then follows immediately that the \(\varphi(r)\) configuration has the following asymptotic behaviors:

\[
\varphi \propto \begin{cases} r & \text{at } r \to 0, \\ \frac{1}{r} & \text{at } r \to \infty. \end{cases} \tag{3.47}
\]

The transition occurs at some intermediate scale that we denote as the Vainshtein radius \(R_V\), within which the nonlinear self-interaction of \(\pi\) becomes important. Furthermore, we can see that when a test particle is placed very close to the source (i.e. \(r \ll R_V\)), the scalar force mediated by the \(\pi\) field is always negligible compared to the normal gravitational force, which goes like \(r^{-1}\). That is the Vainshtein screening mechanism we are after.
It is worth remarking that the Galileon model is a generalization of the 4D effective description of the DGP gravity theory \cite{26, 49, 50}, in the sense that the latter is just a special class of Galileon theory with the coefficients $c_4$, $c_5$ set to zero. The strong coupling scale of the Galileon model was estimated in Ref. \cite{26}. It is much higher than $R_V^{-1}$, implying that our estimate above is still within the valid region of the Galileon theory as an effective theory.

There are more investigations we need to perform on modified gravity models to understand whether they have, besides theoretical interests, practical usefulness. In the next two sections, we will take a closer look at some aspects of the chameleon model and the Galileon model, illustrating the impact on each theory due to observational and theoretical constraints.

### 3.3 No-Go Theorems for Generalized Chameleon Field Theories

We will prove, under very general conditions, two theorems limiting the extent to which chameleon-like theories can impact cosmological observations. The theorems apply to a broad class of these chameleon, symmetron and dilaton theories. The key input is demanding that the Milky Way galaxy, or the Sun, be screened, which is a necessary condition to satisfy local tests of gravity.

The first theorem is an upper bound on the chameleon Compton wavelength at present cosmological density:

$$m_0^{-1} \lesssim \text{Mpc}.$$  \hspace{1cm} (3.48)

Since the chameleon force is Yukawa-suppressed on scales larger than $m_0^{-1}$, this implies that its effects on the large scale structures are restricted to non-linear scales. Any cosmological observable probing linear scales, such as redshift-space distortions, should therefore see no deviation from general relativity in these theories. While the bound (3.48) also appeared independently in \cite{52}, the proof presented here follows a different approach.

The second theorem pertains to the possibility of *self-acceleration*. In the context of chameleon-like theories, by self-acceleration we mean accelerated expansion in the Jordan
frame, while the Einstein-frame expansion rate is not accelerating. This is a sensible definition, for the lack of acceleration in Einstein frame — where the Einstein, and therefore the standard Friedmann, equations hold — is equivalent to the lack of dark energy. In self-accelerating theories, the observed (Jordan-frame) cosmic acceleration stems entirely from the conformal transformation (3.33), i.e., a genuine modified gravity effect. As we showed in the last subsection, the Galileon theory provides such an example among its solutions, where the Einstein frame metric is Minkowski and the Jordan frame metric is de Sitter. However, since the chameleon mechanism relies crucially on the scalar field’s coupling to matter, in chameleon-like theories, we cannot study self-accelerating solution in an empty space (i.e. a space with vanishing $T^{\mu\nu}_{M}$).

Clearly a necessary condition for self-acceleration is that the conformal factor $A(\phi)$ varies by at least $O(1)$ over the last Hubble time $^2$. We will instead find for chameleon-like theories

$$\frac{\Delta A}{A} \ll 1,$$

(3.49)

ruling out the possibility of self-acceleration. Jordan- and Einstein-frame metrics are indistinguishable, and cosmic acceleration requires a negative-pressure component.

Taken together, (3.48) and (3.49) imply that chameleon-like scalar fields have a negligible effect on density perturbations on linear scales, and cannot account for the observed cosmic acceleration except as some form of dark energy. This applies to a broad class of chameleon, symmetron and dilaton theories, including the popular example of $f(R)$. In other words, any such model that purports to explain the observed cosmic acceleration, and passes solar system tests, must be doing so using some form of quintessence or vacuum energy; the modification of gravity has nothing to do with the acceleration phenomenon. Nonetheless, the generalized chameleon mechanism remains interesting as a way to hide light scalars suggested by fundamental theories. The way to test these theories is to study small scale phenomena. Astrophysically, chameleon scalars affect the internal dynamics [53,54] and

$^2$ Relating the Jordan and Einstein frame scale factors by $a_J = A a_E$, it is straightforward to show $[a\ddot{a}]_J - [a\ddot{a}]_E = (A'/A)'$, where $\dot{}$ denotes derivative with respect to (Jordan or Einstein frame) proper time, and $'$ denotes derivative with respect to conformal time. Thus, if $[a\ddot{a}]_E \leq 0$, we must have $[a\ddot{a}]_J \leq (A'/A)'$, implying $1 \lesssim \Delta A/A$ over a (Jordan frame) Hubble time.
stellar evolution \[55-57\] in dwarf galaxies in void or mildly overdense regions.

### 3.3.1 Set-Up

Consider a general scalar-tensor theory in the Einstein frame:

\[
S = \int d^4x \sqrt{-g^E} \left( \frac{R^E}{16\pi G_N} - \frac{1}{2} (\partial \phi)^2 - V(\phi) \right) + S_m[g^J].
\]  

(3.50)

Matter fields described by \(S_m\) couple to \(\phi\) through the conformal factor (positive) \(A(\phi)\) implicit in \(g^J_{\mu\nu}\). The acceleration of a test particle is influenced by the scalar:

\[
\ddot{\mathbf{a}} = -\nabla \Phi_N - \frac{d \ln A(\phi)}{d \phi} \nabla \phi = -\nabla \left( \Phi_N + \ln A(\phi) \right),
\]

(3.51)

where \(\Phi_N\) is the (Einstein frame) Newtonian potential. The fields \(\Phi_N\) and \(\phi\) obey:

\[
\nabla^2 \Phi_N = 4\pi G_N A \rho;
\]

\[
\Box \phi = V_{,\phi} + A_{,\phi} \rho,
\]

(3.52)

where the matter is assumed to be non-relativistic, and \(\rho\) is related to the Einstein and Jordan frame matter densities by \(\rho = \rho_E / A = A^3 \rho_J\) — defined such that \(\rho\) is conserved in the usual sense in Einstein frame. An alternative form of the \(\phi\) equation of motion is useful for comparing against the Poisson equation for \(\Phi_N\):

\[
\Box \phi = 8\pi G_N (V_{,\phi} + \alpha A \rho);
\]

\[
\alpha \equiv \frac{d \ln A}{d \phi} = \frac{M_{Pl}}{\rho},
\]

(3.53)

where \(\varphi \equiv \phi / M_{Pl}\), \(M_{Pl} \equiv (8\pi G_N)^{-1/2}\), and \(\alpha\) quantifies the dimensionless scalar-matter coupling, with \(\alpha \sim \mathcal{O}(1)\) meaning gravitational strength.

A scalar solution of interest is one where \(\phi\) takes the equilibrium value: \(V_{,\phi} + A_{,\phi} \rho = 0\), i.e., \(\rho\) varies sufficiently slowly with space and time such that gradients of \(\phi\) can be neglected. An example is cosmology, with the cosmic mean \(\phi\) adiabatically tracking the minimum \(\phi_{\text{min}}\) of the effective potential \(V_{\text{eff}}(\phi) \equiv V(\phi) + A(\phi) \rho\) as the universe evolves. For simplicity, we assume this minimum is unique, within the field range of interest\(^3\). Further, it is assumed \(\phi_{\text{min}}\) varies monotonically with \(\rho\), say, \(d \phi_{\text{min}} / d \rho \leq 0\) — this is useful for implementing

\(^3\)Symmetrons seem to violate this assumption, by having more than one minimum: different parts of the universe might inhabit different domains. This case is nonetheless covered by our arguments, as long as one interprets the ‘field range of interest’ as that within one domain.
Differentiating $V_{,\phi} + A_{,\phi} \rho = 0$ with respect to $\phi_{\text{min}}$, it is straightforward to show that
\[
\frac{d\phi_{\text{min}}}{d\rho} = -\frac{A_{,\phi}(\phi_{\text{min}})}{m^2},
\]
where
\[
m^2 \equiv V_{\text{eff},\phi\phi}(\phi_{\text{min}}) = V_{,\phi\phi}(\phi_{\text{min}}) + A_{,\phi\phi}(\phi_{\text{min}}) \rho
\]
is assumed non-negative for stability. This means $A$ must be monotonically increasing — hence $V$ must be monotonically decreasing — with $\phi$, at least over the field range of interest. A corollary is that $V(\phi_{\text{min}}(\rho))$ and $A(\phi_{\text{min}}(\rho))$ are respectively monotonically increasing and decreasing functions of $\rho$.

We are particularly interested in the equilibrium $\phi_{\text{min}}$ at cosmic mean density between redshifts $z = 0$ and $z \simeq 1$, the period during which the observed cosmic acceleration commences. Let us refer to the respective equilibrium values: $\phi_{z=0}$ and $\phi_{z=1}$. We are interested in theories with interesting levels of modified gravity effects during this period; we therefore assume:
\[
\alpha(\phi) \gtrsim O(1) \text{ for } \phi_{z=1} \leq \phi \leq \phi_{z=0}.
\]
Note that our set-up automatically guarantees $\phi_{z=1} \leq \phi_{z=0}$. Hence $A(\phi)$ grows with time, which is a necessary condition for self-acceleration.

### 3.3.2 Generalized Screening Condition

Consider a spherically symmetric overdense object that is screened, meaning it sources a scalar force that is everywhere suppressed relative to the gravitational force. According to (3.51),
\[
\frac{d \ln A(\phi)}{dr} \lesssim \frac{d \Phi_N}{dr}.
\]
Both sides of the inequality are positive. The positivity of the right hand side is guaranteed by the positivity of $A\rho$; positivity of the left will be established below. Integrating from inside to outside the object, we have
\[
\ln \left[ \frac{A(\phi_{\text{out}})}{A(\phi_{\text{in}})} \right] \lesssim \Delta \Phi_N.
\]

\footnote{For instance, one would like to demand the scalar mass to be large, or the coupling $\alpha$ to be small, at high $\rho$. But our theorems do not rely on these additional assumptions.}
Here, ‘inside’ means the origin $r = 0$; ‘outside’ means sufficiently far out such that $\phi_{\text{out}}$ is the equilibrium value at today’s cosmic mean density: $\phi_{\text{out}} = \phi_{z=0}$. To satisfy solar system tests, we typically demand that the sun (and also the Milky Way) is screened. Both have a gravitational potential $\Phi_N \sim -10^{-6}$, thus the screening condition is

$$\ln \left[ \frac{A(\phi_{z=0})}{A(\phi_{\text{in-MW}})} \right] \lesssim 10^{-6}.$$  

This inequality will be key in proving (3.48) and (3.49). It makes clear that it is the gravitational potential of the object in question, as opposed to its density alone, that ultimately determines whether it is screened or not.

### 3.3.3 Proof of Theorems

We first rule out self-acceleration by proving (3.49). To do so requires a closer examination of the static and spherically symmetric equation of motion:

$$\phi'' + \frac{2}{r} \phi' = V_{,\phi} + A_{,\phi} \rho,$$  

where $' \equiv \frac{d}{dr}$. This is subject to the boundary conditions $\phi'\big|_{r=0} = 0$ and $\phi \to r \to \infty \phi_{z=0}$. Although $\phi$ tends to its equilibrium value asymptotically, we make no such assumption at the origin, i.e., $\phi\big|_{r=0} \equiv \phi_{\text{in}}$ need not coincide with $\phi_{\text{min}}(\rho_{\text{in}})$. We distinguish 3 cases:

- **Case 1:** Suppose $V_{,\phi} + A_{,\phi} \rho \approx 0$ at $r = 0$, that is, $\phi_{\text{in}} \simeq \phi_{\text{min}}(\rho_{\text{in}})$. This is the thin-shell case of standard chameleons [24]. Since $\rho_{\text{MW}} \gg \rho_{z=1}$, our monotonicity assumptions imply $A(\phi_{z=1}) \geq A(\phi_{\text{in-MW}})$, thus

$$\ln \left[ \frac{A(\phi_{z=0})}{A(\phi_{z=1})} \right] \leq \ln \left[ \frac{A(\phi_{z=0})}{A(\phi_{\text{in-MW}})} \right] \leq 10^{-6}.$$  

This proves (3.49) in this case.

- **Case 2:** Suppose $A_{,\phi} \rho \gg -V_{,\phi}$ at $r = 0$, which is the case relevant to symmetrons [58]. Given our assumption that $V_{\text{eff}} = V(\phi) + A(\phi) \rho$ has a unique minimum, this implies $\phi_{\text{in}} \geq \phi_{\text{min}}(\rho_{\text{in}})$. Because $\phi'\big|_{r=0} = 0$, it follows from (3.59) that $\phi''\big|_{r=0} > 0$, and thus $\phi'\big|_{r>0} > 0$.  

---

5 The screening of the Sun and the Milky Way galaxy go hand-in-hand only because their Newtonian potential happens to be comparable, and for no other reason.
And since $\phi'$ is continuous at the surface of the object, to satisfy $\phi \to r \to \infty \phi_{z=0}$ we must therefore have $\phi_{i_n} < \phi_{z=0}$. In other words, Case 2 corresponds to

$$\phi_{\min}(\rho_{i_n}) \leq \phi_{i_n} < \phi_{z=0}.$$  \hfill (3.61)

Unlike Case 1, $\phi_{i_n-MW}$ is not \textit{a priori} constrained to be smaller (or greater) than $\phi_{z=1}$. If $\phi_{z=1} \geq \phi_{i_n-MW}$, then as in Case 1 we are led to (3.60), and self-acceleration is ruled out. The other possibility, $\phi_{z=1} < \phi_{i_n-MW}$, is inconsistent with screening the Milky Way. Indeed, in this case $\phi$ falls within the range (3.55) where $\alpha(\phi) \sim \mathcal{O}(1)$, and (3.53) can be approximated by $\nabla^2 \varphi \sim 8\pi G N A \rho$. Comparing with the Poisson equation $\nabla^2 \Phi_N = 4\pi G N A \rho$, it is clear the resulting scalar force is not small compared to the gravitational force, thus invalidating the screening of the Milky Way.

- \textbf{Case 3}: Suppose $A,\varphi, \rho \ll -V, \varphi$ at $r = 0$, that is, $\phi_{i_n} \leq \phi_{\min}(\rho_{i_n})$. In this case, all inequalities are reversed relative to Case 2, and instead of (3.61) we conclude $\phi_{\min}(\rho_{i_n}) \geq \phi_{i_n} > \phi_{z=0}$. But this is inconsistent with our assumption that $\phi_{\min}(\rho)$ is monotonically decreasing, hence we can ignore this case.

To summarize, the only phenomenologically viable possibilities are Case 1 and Case 2 with $\phi_{z=1} \geq \phi_{i_n-MW}$. In both cases we are led to (3.60). The very small $\Delta A/A$ over cosmological time scales precludes self-acceleration.

To establish the bound (3.48) on $m_0^{-1}$, consider the (Einstein-frame) cosmological evolution equation:

$$\ddot{\varphi} + 3H \dot{\varphi} = -V, \varphi - A, \varphi \rho,$$  \hfill (3.62)

where $\rho$ is the total (dark matter plus baryonic) non-relativistic matter component, and $H \equiv \dot{a}_E/a_E$ is the Einstein-frame Hubble parameter. Since $A(\phi)\rho \sim H^2 M^2_{Pl}$ from the Friedmann equation, the density term in (3.62) exerts a significant pull on $\phi(t)$. The potential prevents a rapid roll-off of $\phi$ by canceling the density term to good accuracy: $V, \varphi \simeq -A, \varphi \rho$. This cancellation must be effective over at least the last Hubble time, \textit{i.e.} $\phi$ must track adiabatically the minimum of the effective potential. Differentiating this relation with respect to time, and using (3.54) together with the conservation law $\dot{\rho} = -3H \rho$ and the Friedmann relation, we find

$$m^2 \simeq \frac{3HA, \varphi \rho}{\dot{\varphi}} \sim H \frac{dt}{d \ln A} \alpha^2(\phi) H^2.$$  \hfill (3.63)
The factor of $H^{-1} \ln A/dt$ is the change of $\ln A$ over the last Hubble time, which from (3.60) is less than $10^{-6}$. Thus

$$m^2 \gtrsim 10^6 \alpha^2(\phi) H^2.$$  \hfill (3.64)

Using (3.55), it follows that $m_0^{-1} \lesssim 10^{-3} H_0^{-1} \sim \text{Mpc}$, as we wanted to show.

Let us close this subsection with an observation on how theories that screen by the Vainshtein mechanism circumvent our no-go theorems. They replace the potential $V(\phi)$ by derivative interactions. A key effect is that the screening condition (3.56) needs only hold up to some radius, the so-called Vainshtein radius, of the object, thus decoupling $\phi_{\text{out}}$ from $\phi_{z=0}$. It would be interesting to investigate whether chameleon-like theories can also achieve such decoupling.

### 3.4 Classical Stability of Galileon

We begin by assuming some solution, say $\pi_0$, for the Galileon field that satisfies the equations of motion in the presence of some source (3.44), and then expand around this solution by adding small fluctuations. Taking $\pi_0 \to \pi_0 + \phi$ and keeping second order in $\phi$ terms in the Lagrangian, we have the dynamics of the fluctuations. The action is given by

$$S_\phi = \frac{1}{2} \int d^4 x Z_{\mu\nu}(x) \partial_\mu \phi \partial_\nu \phi$$  \hfill (3.65)

where $Z_{\mu\nu}$ is a matrix made out of the second derivatives of our $\pi_0$ field (i.e. made out of the matrix $\partial_\mu \partial_\nu \pi$). \footnote{The reason why only terms proportional to $\partial^2 \phi$ survive in (3.66) is as follows: Note that the $n$th order Lagrangian can be written as $\mathcal{L}_n = T_{\mu_1 \nu_1 \ldots \mu_n \nu_n} \partial_{\mu_1} \pi \partial_{\nu_1} \pi \partial_{\mu_2} \pi \ldots \partial_{\mu_n} \pi$, where $T$ is antisymmetric under changes of any $(\mu_i, \mu_j)$ or $(\nu_i, \nu_j)$ pair while symmetric under that of $(\mu_i, \nu_i)$ pair \footnote{Thus, thanks to the properties $T$ possesses, all the terms in the second variation of $\mathcal{L}_n$ can be expressed as $\partial^2 \phi (\partial \pi_0)^{n-2}$ + surface terms}. Equivalently the matrix $Z_{\mu\nu}$ can be found by taking the first order variation of the equation of motion

$$\delta \mathcal{E}_n = -Z_{\mu\nu} \partial_\mu \partial_\nu \phi$$  \hfill (3.66)

which provides an easy way to compute $Z$. \footnote{The explicit form of various $Z_{\mu\nu}$'s are given.}

Note that this is consistent with (3.65) because $\partial_\mu Z_{\mu\nu} = 0$ identically.
Simply put, the stability conditions for $\pi_0$ configuration are those under which these small fluctuations remain small.

### 3.4.1 Conditions for Stability

The main goal of this subsection is to explore the parameter space spanned by the $c_i$’s ($i = 1, \ldots, 5$) to determine if there exists a subspace in which the Galileon theory is stable. We should point out that the “stability” we refer to is consistent with the physical meaning meant by [50]. While there are other definitions of stability, we utilize this definition because, though it may be limited, it is precise. Given that it is slightly different than the usual definition of stability in ODEs or PDEs, it is worth a few explanations. First, the stability we consider here is a *local* one, i.e., on a space and time scale much shorter than those typical of the background field $\pi_0$, and thus we are safe to treat the matrix function $Z_{\mu\nu}(x)$ as constant. Therefore (it should be noted) we will not be able to keep track of phenomena like resonances which can be interpreted as “instabilities”. These “instabilities” are of a much less catastrophic nature than the ones we discuss and still allow for much analytic control. In the neighborhood of a given point, our stability corresponds to demanding the e.o.m. of the fluctuation field

$$ Z_{\mu\nu} \partial^\mu \partial^\nu \phi = 0 $$ (3.67)

give an oscillating solution. When cast in Fourier space, oscillating solutions correspond to

$$ (Z_{00}\omega + Z_{0i}q^i)^2 = (Z_{0i}Z_{0j} - Z_{00}Z_{ij})q^i q^j $$ (3.68)

having real solutions for $\omega$ for any real spatial momentum vector $\vec{q}$, or equivalently, the matrix $Z_{0i}Z_{0j} - Z_{00}Z_{ij}$ will be positive definite. Second, we want the Galileon theory to be free of ghost-like instability as it is coupled to matter sectors [50][51], i.e. the sign of the kinetic term of the fluctuation action (3.65) to be correct, which requires $Z_{00} > 0$.

As previously mentioned, we will focus on non-relativistic matter sources, not only because of their great importance, but because it can be shown that, given such sources at a generic spacetime point, the symmetric tensor $\Pi_{\mu\nu} \equiv \partial_\mu \partial_\nu \pi$, and therefore $Z_{\mu\nu}$, can be diagonalized through a Lorentz transformation [50]. Thus, the conditions for the local
stability are simply
\[ Z_\mu < 0 , \] (3.69)
where \( Z_\mu \)'s are the diagonal elements of the matrix \( Z_\mu^{\nu} \equiv \text{diag}(Z_0, Z_1, Z_2, Z_3) \).

It has been shown in the DGP\(^8\) model that, given positive energy density sources, if a specific solution is stable at some point (in the way indicated above) then its stability throughout the spacetime is assured [50]. This is, of course, a desirable property of a theory. One may wonder whether a generic Galileon theory shares this same nice property. Are there “safe” choices of the \( c_i \) parameters (evidently, they should include the DGP parametrization) such that this subclass of Galileon theories possess the same property as the DGP model? The answer to this question is the main result of this subsection.

In order to be absolutely clear about what we accomplish, we define the concept of \textit{absolute stability}. Assume some solution to the equations of motion, \( \pi_0 \), exists.

**Definition.** An \textit{absolutely stable region} in parameter space is a region of \{\( c_1, \ldots, c_5 \)\}'s where, if at a single point in spacetime, say \( x_0^{\mu} \), \( Z_\mu(\pi_0(x_0^{\mu})) < 0 \) (i.e. \( \pi_0 \) is stable at this point), then for non-relativistic source profiles satisfying \( \rho \in [0, \infty) \) the equations of motion guarantee that \( Z_\mu < 0 \) over the rest of spacetime (i.e. \( \pi_0 \) is stable over all of space).

Why is this a useful concept? If a choice of parameters is absolutely stable, then it follows that for any non-relativistic positive energy source configuration—no matter what the global structure —stability of a particular solution at a single point implies global stability. When talking about absolute stability one does not have to solve (or at least characterize the solutions) the Cauchy problem for all possible source configurations. We consider the whole equation of motion surface\(^9\) rather than characterize particular solutions.

This stronger cut on acceptable parameters allows us to side-step the difficulties of dealing with the arbitrarily complicated global structure of our source. Considering that we are

---

\(^8\)Throughout this subsection, by DGP model we refer to its 4D effective description rather than the full 5D brane gravity theory.

\(^9\)When we formulate our analysis in terms of the eigenvalues of the \( \Pi_{\mu \nu} \) tensor, the equation of motion \( E = \rho/M_{Pl}^2 \) (at a single point in spacetime, or equivalently, with \( \rho \) fixed) defines a surface in the space spanned by these eigenvalues. It is this surface that we are referring to. All the surfaces generated by different sources (\( \rho \in [0, \infty) \) ) we group into a “family”.

dealing with a non-linear PDE it is surprising that we can say anything at all. In the general case we don’t know how to show existence, but we are still able to say something about stability.\footnote{It should be noted that “absolute stability” does not imply that all solutions are stable, as there can be different branches in our equation of motion surface. But, once again, it does imply that if a solution is stable at one spacetime point it is stable over the rest of spacetime.}

The DGP theory is absolutely stable\footnote{50}. Spherically symmetric solutions (and mild deformations of them) for particular choices of parameters in a single field Galileon model\footnote{26} are stable. Is there some part of this parameter space that admits absolute stability?

To our surprise, we find that the DGP model is the single absolutely stable class. That is, the powerful property of absolute stability that the classical DGP theory possesses does not carry over into the general Galileon theory.

### 3.4.2 General Program

After diagonalizing $\Pi_{\mu\nu}$ with an appropriate boost we can write the stability conditions in a nice algebraic way:

\begin{align*}
Z_0^0 & \equiv Z_0(c_2, c_3, c_4, c_5, k_1, k_2, k_3) < 0 \quad (3.70) \\
Z_1^1 & \equiv Z_1(c_2, c_3, c_4, k_2, k_3) < 0 \quad (3.71) \\
Z_2^2 & \equiv Z_2(c_2, c_3, c_4, k_1, k_3) < 0 \quad (3.72) \\
Z_3^3 & \equiv Z_3(c_2, c_3, k_1, k_2) < 0 \quad (3.73)
\end{align*}

where the $c_i$’s are the coefficients that describe our freedom in choosing the exact Galileon Lagrangain\footnote{3.43} and the $k$’s are the eigenvalues of $\Pi_{\mu\nu}$ for non-relativistic sources. The reason $k_0 (\equiv k_{00} \text{ after the matrix has been diagonalized})$ does not appear in the expression above is because it is suppressed by two powers of $v \ll 1$ in comparison to the other eigenvalues, that is $k_{00} \sim v^2 k_{ij}$, which must be small by assumption in order to ensure diagonalization. $Z_0$ is a cubic function of the $k$’s while the $Z_i$’s are quadratic.

Additionally, we have the equation of motion for $\pi_0$ which becomes an algebraic equation for the $k$’s. Note that $c_1$ and $\rho$ enter in the same manner and are easily combined when we
consider a single point in spacetime. We have

$$\mathcal{E}(c_1, c_2, c_3, c_4, c_5, k_1, k_2, k_3) = \frac{\rho}{M_{Pl}^2}$$

(3.74)

$\mathcal{E}$ is cubic in the $k$’s.

A particular choice of $c_i$’s define a given Galileon theory. They are considered constant over all of spacetime while the values of $\rho(x)$ that characterize our source will have various profiles for different physical configurations. Consider a single point in space, say $\vec{x}_1$.

In eigenvalue space (what we will call “$k$” space) the equations of motion generate a surface (or more correctly surfaces—branches—as our equation is a cubic polynomial) which depend on $\rho(\vec{x})$. The $Z_\mu < 0$ inequalities will define volumes in $k$ space; they are independent of $\rho(\vec{x})$. Say we restrict ourselves to a particular surface generated by $\mathcal{E} = \rho/M_{Pl}^2$.

The question is: given particular values of the $c_i$’s and $\rho(\vec{x}_1)$ are there some values of the $k$’s that lie on this surface but violate $Z_0 < 0$ or $Z_1 < 0$ (it is enough to consider $Z_0$ and a single $Z_i$)? If this is the case, then either the entire surface is embedded inside a region where at least one $Z_\mu > 0$, or this surface intersects with the ‘marginality surfaces’—the surfaces generated by $Z_\mu = 0$.\footnote{Just for emphasis: the space that all these surfaces live in is, of course, the $(k_1, k_2, k_3)$ space. The only function (at this point) of the real space, $(x, y, z)$, is to give us the single point $\vec{x}_1$ whose purpose is to pick out a value of our source, $\rho(\vec{x}_1)$.} If this particular surface of solutions fails to intersect with any of these marginal surfaces, and has at least one particular choice of $k$’s where the stability inequalities hold, then we say that this surface is a 	extit{stable surface} at the point $\vec{x}_1$.

Repeating this analysis at a different point in space, say $\vec{x}_2$, means only taking $\rho(\vec{x}_1) \rightarrow \rho(\vec{x}_2)$. Beyond restricting the sources to positive ones, $\rho(\vec{x}) \geq 0$, a priori we have no idea what the source profile, $\rho(\vec{x})$, will be. Thus, in order to ensure stability of a solution generated by a particular source configuration $\rho(\vec{x})$, we want to find the $c_i$’s such that the family of e.o.m. surfaces generated by the possible values of $\rho(\vec{x})$ is fully embedded within the stable regions ($Z_\mu < 0$). That is, the possible e.o.m. surface at any spatial point is a stable surface.

Usually, the e.o.m. surface $\mathcal{E} = \rho/M_{Pl}^2$ will have multiple branches. A particular solution at one point in real space — i.e. $\pi_0(\vec{x}_0)$ — corresponds to a point in $k$ space sitting on a
particular branch, say Branch 1, of the e.o.m. surface $E = \rho(x_0)/M_{Pl}^2$. If our solution $\pi_0(x)$ is continuous in real space, then $\pi_0(x_0 + \delta x)$ will be on Branch 1 of the e.o.m. surface $E = \rho(x_0 + \delta x)/M_{Pl}^2$. Solutions are confined to the family of a single branch (generated by the possible values of $\rho(x)$) —they cannot jump from one to another. We can therefore analyze each branch in isolation.

At the technical level, we will consider the intersections with the $E = \rho/M_{Pl}^2$ surface with marginal surfaces in k space, which are defined by $Z_\mu = 0$. If it is possible that no intersection occurs, then we must check which side (the stable or unstable) the particular surface generated by the equations of motion falls. In total, consideration of each (no-)intersection will generate a set of ‘stable’ (in the sense given above) $c_i$'s for each marginal surface—if they exist. What are contained in all these sets for all positive values of $\rho$ will constitute the absolutely stable values of the $c_i$'s for the single field Galileon.

In summary, searching for absolute stability corresponds to looking for values of the $c_i$'s (if any) where some particular branch of any e.o.m surface $E = \rho/M_{Pl}^2$ does not inhabit any of the volume excluded by the stability inequalities regardless of the value of $\rho$. Given that the entire e.o.m. surface steers clear of regions of instability, we are assured that if there exists a solution whose value at one point in $x$ space happens to be associated with some point on the stable branch of the e.o.m. surface in k space, then this solution is stable over all of space. Computationally, we determine the absolutely stable region by taking the following steps: 1) Find the conditions for $c_i$’s such that there are no real solutions to the algebraic equations $Z_\mu(k_1, k_2, k_3) = 0$ and $E(k_1, k_2, k_3) = \rho/M_{Pl}^2$ (for a fixed $\rho$). And denote by $V_\rho$ the set of $c_i$’s that satisfy these conditions; 2) Check that some point on the $E(k_1, k_2, k_3) = \rho/M_{Pl}^2$ surface satisfies $Z_\mu < 0$; 3) Repeat this process for all $\rho \in [0, +\infty)$ to obtain the absolute stable region by intersecting all $V_\rho$ over all positive $\rho$.

3.4.3 Details of Analysis

As previously mentioned, in the presence of non-relativistic sources we can diagonalize the matrix $\partial_\alpha \partial_\beta \pi_0$ at a point by an appropriate Lorentz transformation. We can then write the

\[\text{As we will see, in practice, it suffices to consider only two of them: } Z_0 = 0 \text{ and any of } Z_i = 0.\]
CHAPTER 3. EFT APPROACH TO GRAVITY THEORIES: A MODIFICATION OF EINSTEIN GRAVITY

Z’s, (3.65), and the equation of motion of \( \pi_0 \), (3.44), as

\[
Z_\mu(\pi_0) = - \left[ c_2 + 2c_3 \left( \sum_{\alpha=0}^{3} k_\alpha \right) - k_\mu \right] - 6c_4 \left[ \frac{1}{2} \sum_{\alpha \neq \beta} k_\alpha k_\beta \right] - 3 \sum_{\alpha} k_\alpha k_\mu \\
+ 24c_5 \left[ \frac{k_0 k_1 k_2 k_3}{k_\mu} \right]
\] (3.75)

\[
E = c_1 + c_2 \sum_{\alpha} k_\alpha + c_3 \sum_{\alpha \neq \beta} k_\alpha k_\beta + c_4 \sum_{\alpha \neq \beta \neq \gamma} k_\alpha k_\beta k_\gamma \\
+ c_5 \sum_{\alpha \neq \beta \neq \gamma \neq \delta} k_\alpha k_\beta k_\gamma k_\delta - \frac{\rho}{M_{pl}^2} = 0
\] (3.76)

where \( \partial_{\alpha} \partial_{\beta} \pi_0 = \Pi_{\alpha\beta} = \text{diag}(k_0, k_1, k_2, k_3) \). As mentioned above, classical stability corresponds to \( Z_0, Z_1, Z_2, Z_3 < 0 \).

Since we have a non-relativistic source we can consistently suppress the \( k_0 \) dependence in the above expressions as it is suppressed by \( v^2 \) and recover the static limit. In particular

\[
E = c_1 + c_2 (k_1 + k_2 + k_3) + 2c_3 (k_1 k_2 + k_1 k_3 + k_2 k_3) + 6c_4 (k_1 k_2 k_3) - \frac{\rho}{M_{pl}^2}
\] (3.77)

and

\[
Z_0 = - [c_2 + 2c_3 (k_1 + k_2 + k_3) + 6c_4 (k_1 k_2 + k_2 k_3 + k_1 k_3) + 24c_5 k_1 k_2 k_3] \quad (3.78)
\]

\[
Z_1 = - [c_2 + 2c_3 (k_2 + k_3) + 6c_4 (k_2 k_3)] \quad (3.79)
\]

and similarly for \( Z_2 \) and \( Z_3 \).

In the rest of this subsection, we will proceed by analyzing the various possible scenarios—the \( E = \rho/M_{pl}^2 \) surface intersecting with the \( Z_0 = 0 \) or \( Z_1 = 0 \) surfaces—independently, and then consider the intersection of their constraints. To claim absolute stability we must then further take the intersection of the combined constraints for all positive values of \( \rho \). We will concentrate on cases with a nonvanishing \( c_4 \). The special cases associated with \( c_4 = 0 \) are discussed in [28].

An important point which was also shown in [28]: if one branch of the equation of motion surface intersects a marginal surface, then all others do as well. That is, it is enough to find a single intersection for a particular set of \( c_i \)’s to rule out absolute stability for that set.
3.4.4 Conditions for intersection of the $\mathcal{E} = \rho/M_{\text{pl}}^2$ and $Z_0 = 0$ surfaces ($c_4 \neq 0$)

Assume for the moment that we are in the stable region of the $Z_i$’s, that is $Z_1 < 0$, $Z_2 < 0$, and $Z_3 < 0$. We can express $k_1$, $k_2$, and $k_3$ in terms of the particular values of the $Z_i$’s. These are

$$k_i = -\frac{c_3}{3c_4} \pm \frac{\sqrt{f_1f_2f_3}}{6c_4^{1/2}f_i},$$

where

$$f_i \equiv f(Z_i) = -3Z_i + \frac{2c_2^2}{c_4} - 3c_2 \geq \frac{2c_3^2}{c_4} - 3c_2 \geq 0$$

The last inequality comes from constraints obtained by analyzing the stability and existence of radial solutions of the Galileon theory [26]

We normalize the field such that it doesn’t carry any dimensions, that is $[\pi] = M^0 = 1$ where the brackets mean the usual “dimensions of” and $M$ means ”dimensions of mass”. As the action is dimensionless in natural units we have that

$$[c_n] = [M^2]^{2-n}$$

In order to compare the free parameters in the Lagrangian we define dimensionless quantities from the dimensionful $c_i$’s

$$\alpha_1 \equiv \left( c_1 - \frac{\rho}{M_{\text{pl}}^2} \right) \sqrt{c_4}, \quad \alpha_2 \equiv c_2, \quad \alpha_3 \equiv \frac{c_3}{\sqrt{c_4}}, \quad \alpha_5 \equiv \frac{c_5}{c_4^{3/2}}$$

From (3.82), it immediately follows that

$$\alpha_3 \geq \sqrt{\frac{3}{2}} \alpha_2, \quad \text{and} \quad \alpha_5 < 0$$

Note that while originally we had the whole set of parameters $\{c_1, c_2, c_3, c_4, c_5\}$, a stable choice of the these parameters depends only on the choice of the dimensionless parameters...
{α₁, α₂, α₃, α₅}. The c₄ dependence disappears because we use it as our units—we measure everything in terms of c₄. [13] At the classical level, provided it has the correct sign, the overall normalization of the fields does not matter. Noting that α₂ = c₂ > 0, we can therefore work with \(\{\frac{α₁}{α₂}, \frac{α₃}{α₂}, \frac{α₅}{α₂}\}\) which for brevity of notation we define as \(\{α₁, α₃, α₅\}\). Equivalently, we are free to normalize our Lagrangian with the simple choice of \(c₂ = α₂ = 1\).

Either way, we are left with three parameters \(\{α₁, α₃, α₅\}\). We are now ready to investigate the intersection conditions.

Inserting our solutions for the \(kᵢ\)'s, (3.80) (3.81), into \(Z₀\) and the equation of motion, we have

\[
Z₀ = -c₂ + \frac{8α₃α₅}{9} \left( \frac{1}{3} - \frac{4α₃α₅}{9} \right) t \mp \left( \frac{2\sqrt{2}α₅}{9} \right) δ
\]

\[
= (-\frac{\sqrt{2}α₃}{3} + \frac{4\sqrt{2}α₃^2α₅}{9}) uδ = 0
\]

(3.86)

\[
E = α₁ - α₃ + \frac{4α₃^3}{9} ± \frac{δ}{9\sqrt{2}} ± \left( \frac{1}{3\sqrt{2}} - \frac{\sqrt{2}α₃^3}{9} \right) uδ = 0
\]

(3.87)

where

\[
t = f₁ + f₂ + f₃, \quad u = \frac{1}{f₁} + \frac{1}{f₂} + \frac{1}{f₃}, \quad δ = \sqrt{f₁f₂f₃}
\]

(3.88)

Using (3.86) and (3.87) we can always express \(u\) and \(δ\) in terms of \(t\), which we will treat as a free parameter. That is, for some fixed value of \(t\) we can solve for \(u(t)\) and \(δ(t)\) using the above constraints. Furthermore, we can solve for \(f₁, f₂,\) and \(f₃\) via the algebraic equation

\[
F₁(x) \equiv x^3 - tx^2 + u(t)δ(t)^2x - δ(t)^2 = (x - f₁)(x - f₂)(x - f₃) = 0
\]

(3.89)

Finally, we can invert

\[
f(Zᵢ) = -3Zᵢ + 2α₃^2 - 3α₂
\]

(3.90)

to obtain the \(Zᵢ\)'s.

‘Instability’: For a particular choice of the \(cᵢ\)'s or \(αᵢ\)'s such that for some value of \(t\), the \(Zᵢ\)'s are found to be negative implies that a solution could cross into the instability region. Thus, there could exist unstable solutions (solutions in the \(Z₀ > 0\) volume). It

[13] We could have used any other dimensionful parameter, but we find \(c₄\) a convenient choice.
is not guaranteed that the solution is unstable, but rather this is a possibility. Using our terminology: there will be no absolute stability. Hence the quotes.

**Stability:** A stable choice of \(c_i\)'s or \(\alpha_i\)'s corresponds to there being no intersection of the surface of equation of motion and the \(Z_0 > 0\) volume of instability in the \(\{k_1, k_2, k_3\}\) space. Thus, we are seeking a choice of \(\alpha_i\)'s such that, for any \(t\), the equation \(F_t(x) = 0\) cannot have three real roots, all of which must be greater than \((2\alpha_3^2 - 3)\). This can be written in the statement

\[
\{t | t \geq 3(2\alpha_3^2 - 3), \quad \Delta_3[F_t] \geq 0, \quad F_t(2\alpha_3^2 - 3) \leq 0, \quad \text{and } F_t'(x) \geq 0 \text{ for any } x \leq 2\alpha_3^2 - 3\} = \emptyset
\]

where \(\Delta_3[F_t]\) is the discriminant of the cubic equation \(F_t(x) = 0\).

By virtue of the quadratic nature of \(F_t'(x)\), the last condition can be further simplified. Indeed, the axis of symmetry of the upward-opened parabola \(F_t'(x) = 3x^2 - 2tx + u(t)\delta^2(t)\) is \(x = \frac{t}{3} \geq 2\alpha_3^2 - 3\), so the last condition can be replaced by \(F_t'(2\alpha_3^2 - 3) \geq 0\).

Interestingly, one observes that both roots of solution (3.80) yield the same value of \(u(t)\delta^2(t)\) and \(\delta^2(t)\) upon which our auxiliary function \(F_t(x)\) depends. Therefore, both roots actually give the same condition for the stable choice of \(\alpha\)'s and henceforth we can focus on either one.

In summary, the stability condition coming from demanding that the e.o.m. surface *not* intersect the \(Z_0 = 0\) marginal plane reads

\[
\{t | t \geq 3(2\alpha_3^2 - 3), \Delta_3[F_t] \geq 0, F_t(2\alpha_3^2 - 3) \leq 0, \text{and } F_t'(2\alpha_3^2 - 3) \geq 0\} = \emptyset \quad (3.91)
\]

Now, as mentioned in our general outline, we still need to check that the surface generated by the e.o.m. is on the side of stability (\(Z_0 < 0\)) so that we know we are seeing absolute stability as opposed to guaranteed instability. But we hold off on this final check for just a moment.
conditions imply which could be further simplified, when combined with (3.91), to by (3.79) we are free to write the equation of motion as Assume that we are in the \( Z_2 \), \( Z_3 \) and \( Z_0 \) stability region. Using the expression for \( Z_1 \) given by (3.79) we are free to write the equation of motion as

\[
\mathcal{E} = c_1 + c_2(k_2 + k_3) + 2c_3(k_2k_3) - k_1Z_1 - \frac{\rho}{M_{pl}^2} = 0
\]  

(3.92)

\[
\Rightarrow \mathcal{E} = c_1 + c_2(k_2 + k_3) + 2c_3(k_2k_3) - \frac{\rho}{M_{pl}^2} = 0
\]  

(3.93)

Solving the above equation together with the \( Z_1 = 0 \) equation yields the solutions

\[
k_{2,3} = \frac{-3\alpha + \alpha_3 \pm \sqrt{\Gamma} \frac{1}{\sqrt{c_4}}}{6 - 4\alpha_3} \]  

(3.94)

where we have already normalized everything such that \( \alpha_2 = 1 \) and \( \Gamma \) is given by

\[
\Gamma = 9\alpha_1^2 + 6 - 18\alpha_1\alpha_3 - 3\alpha_3^2 + 8\alpha_1\alpha_3^3. \]  

(3.95)

The plus or minus indicated above means that \( k_2 \) must take the plus while \( k_3 \) must take the minus, or vice versa. We pick one. Plugging these solutions into the expressions for \( Z_2 \), \( Z_3 \) and \( Z_0 \), all of which are negative given our assumptions, yields

\[
Z_0 = - \left( \frac{2k_1\sqrt{c_4}}{2\alpha_3^2 - 3} \right) (9\alpha_1 - 6\alpha_3 + 2\alpha_3^2 + 6\alpha_5 - 12\alpha_1\alpha_3\alpha_5) \leq 0
\]  

(3.96)

\[
Z_2 = - \left( \frac{2\alpha_3^2 - 3}{2\alpha_3^2 - 3} \right) (9\alpha_1 - 9\alpha_3 + 4\alpha_3^2)k_1\sqrt{c_4} + (-3 + 3\alpha_1\alpha_3 + \alpha_3^2) + \sqrt{\Gamma}(3k_1\sqrt{c_4} + \alpha_3) \leq 0
\]  

(3.97)

\[
Z_3 = - \left( \frac{2\alpha_3^2 - 3}{2\alpha_3^2 - 3} \right) (9\alpha_1 - 9\alpha_3 + 4\alpha_3^2)k_1\sqrt{c_4} + (-3 + 3\alpha_1\alpha_3 + \alpha_3^2) - \sqrt{\Gamma}(3k_1\sqrt{c_4} + \alpha_3) \leq 0
\]  

(3.98)

The latter two conditions imply

\[
(9\alpha_1 - 9\alpha_3 + 4\alpha_3^2)k_1\sqrt{c_4} + (-3 + 3\alpha_1\alpha_3 + \alpha_3^2) \geq \sqrt{\Gamma|(3k_1\sqrt{c_4} + \alpha_3)|}
\]  

(3.99)

A stable choice of parameters corresponds to

\[
\{k_1\sqrt{c_4} | 9\alpha_1 - 6\alpha_3 + 2\alpha_3^2 + 6\alpha_5 - 12\alpha_1\alpha_3\alpha_5) \geq 0, \ \Gamma \geq 0, \ \text{and} \ \((9\alpha_1 - 9\alpha_3 + 4\alpha_3^2)k_1\sqrt{c_4} + (-3 + 3\alpha_1\alpha_3 + \alpha_3^2)) \geq \sqrt{\Gamma|(3k_1\sqrt{c_4} + \alpha_3)|} \} = \emptyset
\]  

(3.100)

which could be further simplified, when combined with (3.91), to

\[
\alpha_3 - \frac{4}{9}\alpha_3^3 - \frac{\sqrt{\Gamma}}{9}(2\alpha_3^2 - 3)^{3/2} < \alpha_1 < \alpha_3 - \frac{4}{9}\alpha_3^3 + \frac{\sqrt{\Gamma}}{9}(2\alpha_3^2 - 3)^{3/2}
\]  

(3.101)
CHAPTER 3. EFT APPROACH TO GRAVITY THEORIES: A MODIFICATION OF
EINSTEIN GRAVITY

We will leave the lengthy algebraic analysis that generates the above conditions to Appendix D, to which careful readers are referred.

3.4.6 Stable Region: From Local stability to Absolute Stability

We must, of course, take the intersection of the stable choices coming from both conditions, (3.91) and (3.101).

So far, we have worked out the stability conditions for a given choice of $\alpha_i$’s, with the external source $\rho$ being taken as a fixed parameter like the intrinsic ones ($c_i$’s) that define the theory. Say that, at the particular spatial point we are working at, the value of $\rho$ is such that the e.o.m. surface $\mathcal{E} = \rho/M_{Pl}^2$ are completely embedded in the stability region. As we move to a new point in space $\rho$ will generically change. Thus, we need to do our analysis all over again for this new value of the source. The convenience of our method is that the details of how the source changes are washed out as we consider the whole family of surfaces generated by the e.o.m. rather than any particular solution. We lose some information, but we have made the problem tractable. We don’t have to deal with the functional dependence of our source, $\rho(\vec{x})$, boundary conditions, etc. All that matters is the range of values $\rho$ takes, $[\rho_{\text{min}}, \rho_{\text{max}}]$. The values of $\rho$ encountered in any kind of astrophysical/cosmological application of the Galileon theory will be vast, spanning over 40 orders of magnitude from the average density of the universe ($\sim (10^{-3} eV)^4$) to nuclear density ($\sim (GeV)^4$). To describe a universe like our own in our units we can be free to take $\rho_{\text{min}} \to 0$ and $\rho_{\text{max}} \to \infty$.

For a particular value of $\rho$, we are given a particular subset in parameter space $\{c_1, c_3, c_5\}$ (in units of $c_4$) by satisfying (3.91) and (3.101). Let’s call this set of stable parameter choices $A(\rho)$. In order to achieve absolute stability we need to take the intersection of the $A(\rho)$’s of all possible $\rho$’s. Thus

$$\text{Absolutely stable region of } \{c_1, c_3, c_5\}'s = \bigcap_{\rho=0}^{\infty} A(\rho) \quad (3.102)$$

However, it is easily seen that (3.101) cannot hold for any source, since as $\rho \to \infty$, $\alpha_1$ becomes more and more negative (for a given $c_1$) and eventually fails to fall into the region specified by the fixed value of $\alpha_3$ (3.101). Before checking whether the surface generated
by the e.o.m. lies in the stable region, we can draw some conclusions about the absolutely stability region. That is, from the analysis above, we see that for generic sources, no matter how we choose the parameters that define our theory (finite values of $c_1$, $c_3$, and $c_5$) the surface generated by the e.o.m. will pierce the marginal surface generated by $Z_1 = 0$ soiling any hope of absolute stability in our theory.

We can use similar methods to examine the various special situations with $c_4 = 0$, we find that there is no absolutely stable choice of parameters except for the $c_4 = 0, c_5 = 0$ case, which is exactly the DGP model.
Chapter 4

An Inflation Model from EFT Construction: Solid Inflation

4.1 The Effective Theory for a Solid in Flat Spacetime

Like fluids, solids can also be regarded as some space-filling continua, so we expect our effective field theoretic construction to be applicable to solid systems as well. However, in contrast to fluids, solids are sensitive to any sort of displacement, even those preserving the volume of each solid element, due to the presence of anisotropic stresses. This suggests that there is no volume-preserving symmetry for a solid system. In other words, in our effective field theoretic language, fluids are very symmetric solids.

So, the low energy effective action for solids should consist of three scalar fields $\phi^I(x)$ (interpreted as the comoving coordinates as before) obeying Poincaré invariance and the internal symmetries $\text{(2.2)}$ $\text{(2.3)}$. The ground state configuration is once again given by

$$\phi^I = x^I, \quad I = 1, 2, 3. \quad (4.1)$$

At lowest order in the derivative expansion, the only Lorentz-scalar, shift-invariant quantity is the matrix

$$B^{IJ} \equiv \partial_\mu \phi^I \partial^\mu \phi^J. \quad (4.2)$$

$^1$Since we have imposed the full $SO(3)$ invariance, in fact we are describing an isotropic solid with no preferred axes — a ‘jelly’.
We then have to construct $SO(3)$ invariants out of this matrix. For a $3 \times 3$ matrix, there are only three independent ones, which we can take for instance to be the traces

\[ [B], \quad [B^2], \quad [B^3], \quad (4.3) \]

where as before the brackets $[\ldots]$ are shorthand for the trace of the matrix within. Alternatively, one could take the determinant, and two of the traces above. In the following, we will find it convenient to use one invariant—say $[B]$—to keep track of the overall ‘size’ of the matrix $B$, and to choose the other two such that they are insensitive to an overall rescaling of $B$, e.g.

\[ X = [B], \quad Y = \frac{[B^2]}{[B]^2}, \quad Z = \frac{[B^3]}{[B]^3}. \quad (4.4) \]

The most general solid action therefore is

\[ S = \int d^4x F(X,Y,Z) + \ldots \quad (4.5) \]

where $F$ is a generic function that depends on the physical properties of the solid—e.g. its equation of state—and the dots stand for higher-derivative terms, which are negligible at low energies and momenta.

The background configurations $(4.1)$ spontaneously break some of our symmetries. There are associated Goldsone bosons, which are nothing but fluctuations of the $\phi^I$’s about such a background,

\[ \phi^I = x^I + \pi^I. \quad (4.6) \]

We get these fluctuations’ free action by expanding our action $(4.5)$ to second order in $\pi^I$. Using

\[ B^{IJ} = \delta^{IJ} + \partial^I \pi^J + \partial^J \pi^I + \partial_\mu \pi^I \partial^\mu \pi^J, \quad (4.7) \]

after integrating by parts and neglecting boundary terms we get

\[ S \to S_2 = \int d^4x \left[ -\frac{1}{3} F_X X \cdot \vec{\pi}^2 + \left( \frac{1}{3} F_X X + \frac{6}{27} (F_Y + F_Z) \right) (\partial_i \pi_j)^2 \right. \]

\[ + \left. \left( \frac{1}{9} F_{XX} X^2 + \frac{2}{27} (F_Y + F_Z) \right) (\vec{\nabla} \cdot \vec{\pi})^2 \right], \quad (4.8) \]

where the subscripts stand for partial derivatives, which are to be evaluated at the background values

\[ X \to 3, \quad Y \to 1/3, \quad Z \to 1/9. \quad (4.9) \]
These Goldstone excitations are the solid’s phonons. As before it will be convenient to split the phonon field $\vec{\pi}$ into a longitudinal part and a transverse one,

$$\vec{\pi} = \frac{\vec{\nabla}}{\sqrt{-\nabla^2}} \pi_L + \vec{\pi}_T, \quad \vec{\nabla} \cdot \vec{\pi}_T = 0.$$  

(4.10)

It is straightforward to extract the longitudinal and transverse propagation speeds from the phonon’s action:

$$c_L^2 = 1 + \frac{2}{3} \frac{F_{XX} X^2}{F_{XX}}, \quad c_T^2 = 1 + \frac{2}{3} \frac{(F_Y + F_Z)}{F_{XX}},$$  

(4.11)

in terms of which the quadratic action is simply

$$S_2 = \int d^4x \left( -\frac{1}{3} F_{XX} \left[ \dot{\vec{\pi}}^2 - c_T^2 \left( \partial_i \pi_j \right)^2 - (c_L^2 - c_T^2) (\vec{\nabla} \cdot \vec{\pi})^2 \right] \right).$$  

(4.12)

If we expand Eqn. (4.5) to higher orders, we get the interactions among the phonons. The expansion is straightforward, but already at cubic order the result is quite messy, and not particularly illuminating. Roughly speaking, the $n$-th ($n \geq 3$) order interaction terms will be schematically of the form $(\partial \pi)^n$, where the derivatives can be spacial or temporal, and the indices are contracted in all possible ways. The coefficients of these interactions terms—the coupling constants—will be given by suitable derivatives of $F$, evaluated on the background solution. Like for all derivatively-coupled theories, our interactions become strong in the UV, at some energy scale $\Lambda_{\text{strong}}$. For our theory to be predictive for cosmological observables, we will need this strong-coupling scale to be above the Hubble rate $H$, for the whole duration of inflation.

### 4.2 Solids in a Cosmological Background: Inflation

We can now allow for a cosmological spacetime metric and for dynamical gravity, which, operationally, is trivial: the index-contraction in (4.2) should be done via $g^{\mu\nu}$ rather than $\eta^{\mu\nu}$, and the measure in (4.5) should carry a $\sqrt{-g}$. As usual, ‘minimal coupling’ corresponds to the most general coupling one can have between a matter system and gravity at lowest order in the derivative expansion. Then our solid’s stress-energy tensor is

$$T_{\mu\nu} = -\frac{2}{\sqrt{-g}} \frac{\delta S}{\delta g^{\mu\nu}} = -2 \frac{\partial F}{\partial B^{ij}} \partial_\mu \phi^i \partial_\nu \phi^j + g_{\mu\nu} F.$$  

(4.13)
As to the scalar fields' background configuration, the $x^I$ in (4.1) should now be interpreted as *comoving* FRW coordinates. The reason is that the FRW metric is invariant under translations and rotations acting on the comoving coordinates, and we want the l.h.s. and the r.h.s. in (4.1) to transform in the same way under the symmetries we are trying to preserve.

When computed on the background, the stress-energy tensor reduces to the standard $T^\mu_\nu = \text{diag}(-\rho, p, p, p)$, with

$$
\rho = -F, \quad p = F - \frac{2}{a^2} F_X,
$$

where the subscript $X$ stands for partial derivative, and $F$ and $F_X$ are evaluated at the background values for our invariants:

$$
X \rightarrow 3/a^2(t), \quad Y \rightarrow 1/3, \quad Z \rightarrow 1/9.
$$

Notice that—by construction—$X$ is the only invariant that depends on the scale factor; $Y$ and $Z$ were designed to be insensitive to an overall rescaling of $B^{IJ}$. This is the reason why only $F_X$ appears in the pressure: for an FRW solution, the pressure is related to the response of the system to changing the scale factor, i.e., the volume. For a more general configuration, the stress-energy tensor (4.13) has a more complicated structure, which depends on $F_Y$ and $F_Z$ as well, which we report here for later use:

$$
T^\mu_\nu = g^\mu_\nu F - 2 \partial_\mu \phi^I \partial_\nu \phi^J \left( X \frac{2F_Y Y}{X} - 3 \frac{F_Z Z}{X^3} \right) \delta^{IJ} + \frac{2F_Y B^{IJ}}{X^2} + \frac{3F_Z B^{IK} B^{KJ}}{X^3}.
$$

Now, in order to have near exponential inflation, we need

$$
\epsilon \equiv -\frac{\dot{H}}{H^2} \ll 1.
$$

Via the Friedmann equations\(^2\)

$$
H^2 = \frac{1}{3M_{\text{Pl}}^2} \rho, \quad \dot{H} = -\frac{1}{2M_{\text{Pl}}^2} (\rho + p),
$$

and eq. (4.14), we can express $\epsilon$ directly in terms of our Lagrangian $F$:

$$
\epsilon = 3 \cdot \frac{1}{a^2} \frac{F_X}{F} = \frac{\partial \log F}{\partial \log X},
$$

\(^2\)We are defining the Planck scale as $M_{\text{Pl}}^2 = (8\pi G)^{-1}$
where we used eq. (4.15) for the background value of $X$. We thus see that if we want our solid to drive near exponential inflation, we need a very weak $X$-dependence for $F$. Which is not surprising, since $X$ is the only invariant that is sensitive to the volume of the universe: for inflation to happen, the solid’s energy should not change much if we dilate the solid by $\sim e^{60}$; this is only possible if the solid’s dynamics do not depend much on $X$.

This also suggests how to enforce the smallness of $\epsilon$ via an approximate symmetry. Consider the scale transformation

$$\phi^I \rightarrow \lambda \phi^I, \quad \lambda = \text{const}.$$ (4.20)

The matrix $B^{IJ}$ changes by an overall $\lambda^2$ factor, which affects $X$ but not $Y$ nor $Z$. Therefore, the smallness of $F_X$ can be interpreted as an approximate invariance under (4.20): If $F$ only depended on $Y$ and $Z$, it would be exactly invariant under (4.20), and this would prevent quantum corrections from generating some $X$-dependence. If we start with a small $F_X$ at tree level, the symmetry is only approximate, yet all further $X$-dependence generated at quantum level will be suppressed by the small symmetry-breaking coupling constant—$F_X$ itself.

Notice that here we are dealing with a purely internal symmetry—eq. (4.20)—which commutes with all spacetime symmetries. It has nothing to do with spacetime scale-invariance. It is on an equal footing with our other internal symmetries (2.2), (2.3), and, like those, is non-anomalous and can be used to constrain the structure of the Lagrangian. To avoid confusion, in the following we will refer to the symmetry (4.20) as ‘internal scale invariance’.

As manifest from our quadratic action for the phonons—eq. (4.8)—, the phonon’s kinetic energy is suppressed by $\epsilon$,

$$S_2 \sim \int d^4x \epsilon |F| \cdot \dot{\pi}^2 + \ldots$$ (4.21)

For very small $\epsilon$, this can in principle lead to two problems for our theory:

- Superluminality: the gradient energies in (4.8) are not explicitly suppressed by $\epsilon$, and as a consequence the propagation speeds (4.11) are formally of order $1/\epsilon$, unless the numerators are also small. In an effective field theory like ours, with spontaneously
broken Lorentz invariance, superluminal signal propagation is not necessarily an inconsistency. However, it prevents the theory from admitting a standard Lorentz-invariant UV-completion \cite{60}. We therefore feel that it should be avoided.

- **Strong coupling:** unless interactions are also suppressed by suitable powers of $\epsilon$—and it turns that they are not—a smaller kinetic energy means stronger interactions. This is obvious if one goes to canonical normalization for $\pi$, by absorbing the prefactor in (4.21) into a redefined phonon field. Then inverse powers of $\epsilon$ will show up in the interaction terms, thus signaling that the strong coupling scale of the theory is suppressed by some (positive) power of $\epsilon$. We have to make sure that this strong coupling scale is above $H$, for the whole duration of inflation.

As for the former issue, notice first of all that the term proportional to $F_{XX}$ in the expression for $c_L^2$ is forced to be close to $-2/3$. The reason is that not only do we need the ‘slow-roll’ condition (4.17) for inflation to happen, we also need

$$\eta = \frac{\dot{\epsilon}}{\epsilon H} \ll 1 \quad (4.22)$$

for inflation to last many $e$-folds\footnote{In the computations that follow, we will assume all the slow-roll parameters to be of the same order of magnitude.}. This forces the second derivative $F_{XX}$ also to be small. In particular, given eq. (4.19), and

$$H = \frac{d}{dt} \log a = -\frac{1}{2} \frac{d}{dt} \log X \quad (4.23)$$

we get

$$\frac{F_{XX} X^2}{F_X X} = -1 + \epsilon - \frac{1}{2} \eta \quad (4.24)$$

So that, at lowest order in $\epsilon$ and $\eta$, the propagation speeds (4.11) reduce to

$$c_L^2 \simeq \frac{1}{3} + \frac{8}{9} \frac{(F_Y + F_Z)}{F_X X}, \quad c_T^2 = 1 + \frac{2}{3} \frac{(F_Y + F_Z)}{F_X X} \quad (4.25)$$

It is quite interesting that in this limit they depend on exactly the same ($F_Y + F_Z$) combination. As a result, the two speeds are not independent: they are related by \footnote{To all orders in $\epsilon$ and $\eta$, the exact relation is}

$$c_T^2 \simeq \frac{3}{4} (1 + c_L^2) \quad (4.27)$$
We thus see that for both speeds to be sub-luminal, we need

\[ 0 < (F_Y + F_Z) < \frac{3}{8} \epsilon |F| \]  

(4.28)

where the upper bound is imposed to ensure that \( c_L^2 \), \( c_T^2 \) is positive.

To fulfill this condition, we could demand that all derivatives of \( F \) be small,

\[ F_Y \sim F_Z \sim F_X \sim \epsilon F, \]

which corresponds to saying that the bulk of the solid’s energy density and pressure are dominated by a cosmological constant, which does not depend on the fields. Although this is of course a technically natural choice—having a large cosmological constant was never a problem—it is not particularly interesting. It would be more interesting to allow for large derivatives of \( F \),

\[ F_Y, F_Z \sim F \]  

(4.29)

but demand the combination \( F_Y + F_Z \) to vanish. While we have motivated the smallness of \( F_X \) (i.e. \( F_X X = \epsilon F \) ) via an approximate symmetry, we have not been able to find a symmetry that enforces the condition (4.28) while preserving (4.29). We have to take such a condition as an assumption, which might involve some fine tuning, but which is nonetheless consistent and necessary for the consistency of our inflationary solution.

As for the strong coupling issue, we have to estimate the strong coupling scale \( \Lambda_{\text{strong}} \) in our small \( \epsilon \) limit, and make sure that cosmological perturbations are weakly coupled at horizon crossing, that is, at frequencies of order \( H \). Expanding the action (4.5) to all orders in \( \pi \) we get interactions of the form

\[ f_n \cdot (\partial \pi)^n, \]

(4.30)

where \( f_n \) is some typical derivative of \( F \). In our case some combinations of derivatives are small,

\[ F_X X \sim (F_Y + F_Z) \sim \epsilon F, \]  

(4.31)

but we do not expect this to yield a substantial weakening of interactions. For instance, we will see below that in our approximation the coefficient weighing cubic interactions is
$F_Y$, which, as we argued above, can be as large as the background energy density, $F_Y \sim F$. Assuming that $F_Y$ is a good estimate for the coefficients encountered in interaction terms, and assuming for the moment that both $c_L$ and $c_T$ are of order of the speed of light—so that there is no parametric difference between time- and space-derivatives—we can estimate very easily the strong coupling scale: We can go to canonical normalization for the kinetic term

$$L_2 \sim \epsilon F \cdot (\partial \pi)^2 \rightarrow (\partial \pi c)^2,$$

so that the $n$-th order interaction becomes

$$L_n \sim F_Y \cdot (\partial \pi)^n \rightarrow \frac{F_Y}{(\epsilon F)^{n/2}} (\partial \pi c)^n.$$  \hfill (4.33)

This is a dimension-2$n$ interaction, weighed by a scale

$$\Lambda_n \sim \left(\frac{\epsilon^n F^n}{F_Y^2}\right)^{\frac{1}{4n-8}}$$  \hfill (4.34)

(recall that $F$ and $F_Y$ have mass-dimension four.) If $F_Y$ is of the same order as $F$, this is simply

$$\Lambda_n \sim F^{1/4} \cdot \epsilon^{\frac{n}{4n-8}}, \quad (F_Y \sim F),$$  \hfill (4.35)

which, for $n \geq 3$ and $\epsilon \ll 1$, is an increasing function of $n$. The lowest of all such scales—which defines the strong coupling scale of the theory—is thus that associated with $n = 3$:

$$\Lambda_{\text{strong}} = \Lambda_3 \sim F^{1/4} \epsilon^{3/4}, \quad (F_Y \sim F).$$  \hfill (4.36)

i.e. we found that the strong coupling scale is a fractional power of $\epsilon$ smaller than the scale associated with the solid’s energy density.

On the other hand, if the value of $F_Y$ is much smaller than that of $F$, or if we use other coefficients in interaction terms that are parametrically smaller than $F$ to do the estimate, the strong coupling scales we get will be higher than that in 4.36, and hence they are less dangerous.

We can run similar arguments for the case where the longitudinal speed of sound is non-relativistic (note that $c_T$ is always relativistic, since it follows from 4.11, 4.28 that $3/4 < c_T^2 < 1$. ) It is straightforward to check that the strong coupling momentum and energy scales are given by

$$p_{\text{strong}} \sim F^{1/4}(\epsilon^3 c_L^5)^{1/4}, \quad E_{\text{strong}} \sim F^{1/4}(\epsilon^3 c_L^9)^{1/4}, \quad (F_Y \sim F).$$  \hfill (4.37)
As we mentioned above, cosmological perturbation theory is under control only if
\[ E_{\text{strong}} \gg H \, . \] (4.38)
Relating \( H \) and \( F \) via the Friedmann equation, \( H^2 \sim F/M_{\text{Pl}}^2 \), we get a lower bound on the combination \( \epsilon \cdot c_L^3 \),
\[ \epsilon \cdot c_L^3 \gg (H/M_{\text{Pl}})^{2/3} \, . \] (4.39)
In principle our \( H \) can be several orders of magnitudes smaller than the Planck scale, in which case this bound is not particularly restrictive. Still, it is a nontrivial condition for the self-consistency of the perturbative computations we will perform.

A clarification is in order: we have been analyzing the viability of our model focusing on the phonons’ dynamics, neglecting the background spacetime curvature and the phonons’ mixing with gravitational perturbations. Of course this is not entirely correct. However, at energies much bigger than \( H \), or equivalently, for time-scales much shorter than \( H^{-1} \), curvature and mixing have negligible effects, and in first approximation they can be neglected. Our conditions above, (4.28) and (4.39), should then be thought of as necessary and sufficient for our system to be well-behaved in the UV, at very short distances and time scales. Our detailed analysis of cosmological perturbations in Section 4.3 will confirm these results.

We should also point out that although we will be using standard ‘slow-roll’ nomenclature for the conditions (4.17), (4.22) and for the associated perturbative expansion, nothing is ‘rolling’ in our system, slowly or otherwise: our \( \phi^I \) scalars are exactly constant in time. As usual however, the so-called slow-roll expansion really relies on the slowness of certain time-dependent observables like \( H, \dot{H}, \) etc., which are well defined regardless of the presence of a rolling scalar. We will still use ‘slow-roll’ to refer to such a weak time-dependence, hoping that this will not cause confusion. As we emphasized, in our case the physical origin of this slowness is the near independence of the dynamics on \( X \), which is, among our invariants, the only one that depends on time. Besides \( \epsilon \) and \( \eta \), in the following we will need one more slow-roll parameter,
\[ s \equiv \frac{\dot{c}_L}{c_L H} \, , \] (4.40)
which is small, because \( c_L \) depends on time only via the Lagrangian’s \( X \)-dependence.
Finally, we should comment on why we are focusing on a solid rather than on a perfect fluid. First, since eventually we will be interested in quantum mechanical effects—as usual, quantum fluctuations will be the ‘seed’ for cosmological perturbations—we focus on a solid because we do not know yet how to consistently treat the perfect fluid effective theory as a quantum theory. The problem has to do with the transverse excitations, which appear to be strongly coupled at all scales. Second, even forgetting about the transverse excitations and focusing on the longitudinal ones, we would not be able to keep those weakly coupled for many e-folds. As clear from (4.26), to have vanishing $c^2_T$ (which is one of the defining features of a fluid) and small $\epsilon$, we need $\eta \sim -1$. But, by definition, $\eta = \dot{\epsilon}/(H\epsilon)$, so that we need $\epsilon = F_X X/F$ to decrease by an order one factor over one Hubble time, i.e. to decrease like some order-one power of $1/a$. $F$ has to be nearly constant over many e-folds, which means that it is actually the numerator $F_X X$ that is tracking $1/a(t)$. But it is precisely combinations like $F_X X$ that control the strong-coupling scale for longitudinal excitations in a fluid, which means that we cannot have $F_X X$ decrease by exponentially large factors without making the system strongly coupled at frequencies of order $H$ at some point during inflation.

4.3 Cosmological perturbations

The three sections that follow contain a technical analysis of cosmological perturbations. Before skipping directly to Section 4.6, the reader uninterested in the details of the derivations should be aware of our results: the scalar tilt (4.91), the tensor-to-scalar ratio (4.92), the tensor tilt (4.76), and the three-point function of scalar perturbations (4.106) (which is analyzed in some detail in Section 4.6).

As the background stress tensor takes the usual homogeneous and isotropic form represented by $T^\mu_\nu = \text{diag}(-\rho, p, p, p)$, all the interesting repercussions of our peculiar symmetry breaking pattern lie in the dynamics of perturbations around the slow roll background. In order to best isolate the dynamical degrees of freedom of the gravitational field it is most convenient to work in the ADM variables introduced in (2.100) (2.101). For the background FRW metric $N = 1$, $N^i = 0$, and $h_{ij} = a^2(t) \delta_{ij}$. 
Following [13] we can write the action as

\[
S = \int d^4x \sqrt{-h} \left\{ \frac{1}{2} M_{Pl}^2 [R^{(3)} + N^{-2} (E_{ij} E^{ij} - E^2)] + F(X, Y, Z) \right\}
\]

(4.41)

where \( R^{(3)} \) is the 3-dimensional Ricci scalar constructed out of \( h_{ij} \) and \( E_{ij} = N K_{ij} \), with \( K_{ij} \) denoting the extrinsic curvature of equal-time hypersurfaces. The constraint equations given by varying (4.41) with respect to \( N \) and \( N^i \) are:

\[
0 = \frac{1}{2} M_{Pl}^2 \left[ R^{(3)} - N^{-2} (E_{ij} E^{ij} - E^2) \right] + F(X, Y, Z) + N \frac{\partial F(X, Y, Z)}{\partial N}
\]

(4.42)

\[
0 = \frac{1}{2} M_{Pl}^2 \nabla_i \left[ N^{-1} (E^i_j - \delta^i_j E) \right] + N \frac{\partial F(X, Y, Z)}{\partial N^j}.
\]

(4.43)

The derivatives of \( F \) with respect to \( N \) and \( N^j \) can be calculated easily by noting that our \( B^{IJ} \) (and hence \( X, Y, Z \)) can be expressed in ADM variables as

\[
B^{IJ} = -\frac{1}{N^2} (\dot{\phi}^I - N^k \partial_k \phi^J) (\dot{\phi}^J - N^k \partial_k \phi^I) + h^{km} \partial_k \phi^I \partial_m \phi^J.
\]

(4.44)

For the moment we find it more convenient to work in spatially flat slicing gauge (SFSG)—defined in Appendix [E]—where we can write the fluctuations about the FRW background as

\[
\phi^I = x^I + \pi^I, \quad h_{ij} = a(t)^2 \exp(\gamma_{ij}), \quad N = 1 + \delta N,
\]

(4.45)

where \( \gamma_{ij} \) is transverse and traceless, i.e.

\[
\partial_i \gamma_{ij} = \gamma_{ii} = 0.
\]

(4.46)

We can also further split the \( \pi^i \) and \( N^i \) fields in terms of their longitudinal scalar and transverse vector components. We therefore write:

\[
\pi^i = \frac{\partial_i}{\sqrt{-\nabla^2}} \pi_L + \pi^i_T, \quad \text{and} \quad N^i = \frac{\partial_i}{\sqrt{-\nabla^2}} N_L + N^i_T,
\]

(4.47)

where \( \partial_i \pi^i_T = \partial_i N^i_T = 0 \). From now on we will stop differentiating between internal \( I, J, \ldots \) indices and spacial \( i, j, \ldots \) ones. The reason is that of the full original \( SO(3)_{\text{spacetime}} \times SO(3)_{\text{internal}} \) symmetry, only the diagonal combination is preserved by the background \( \phi^I = x^I \). \( \pi^i \) and \( N^i \) both transform as vectors under this unbroken \( SO(3) \), and therefore they carry the same kind of index.
For our purposes here we are interested only in the leading non-gaussian behavior. Barring accidental cancellations, this can be captured by keeping terms that are cubic in the fluctuations. In order to reproduce these terms it turns out to be necessary to only know $N$ and $N^i$ to first order in the fluctuations \[^5\] Due to the homogeneity of the 3D-spatial hypersurfaces, we will work in spatial Fourier space, with our convention defined for any field $\xi(x)$ by:

$$
\xi(t, \vec{x}) = \int_{\vec{k}} e^{i\vec{k}\cdot\vec{x}} \tilde{\xi}(t, \vec{k}), \quad \int_{\vec{k}} \equiv \int \frac{d^3k}{(2\pi)^3} .
$$

To simplify the notation, however, we will drop the twiddle as which field variable we intend will be obvious from the arguments. And so, solving the constraint equations (4.42) and (4.43) to first order in fluctuations we have

$$
\delta N(t, \vec{k}) = -\frac{a^2H^2\dot{\pi}_L - \dot{H}\pi_L / H}{k^2 / 1 - 3H a^2 / k^2}
$$

$$
N_L(t, \vec{k}) = -\frac{3a^2H\dot{\pi}_L / k^2 + \dot{H}\pi_L / H}{1 - 3a^2H / k^2}
$$

$$
N_i^T(t, \vec{k}) = \frac{\dot{\pi}_T^i}{1 - k^2 / 4a^2H}
$$

where the dot denotes a time-derivative.

Now, plugging these solutions back into (4.41) will give us the correct action for the fluctuations up to cubic order. For instance, the trilinear solid action after mixing with gravity is contained in Appendix [F] while the quadratic actions for the tensor, vector and scalar modes are contained in the next section.

Now that we have the correct action for the perturbations in the presence of an inflating background we can compute correlation functions. In the end, we are interested in the post-reheating correlation functions of curvature perturbations, parameterized by either of the gauge invariant (at linear-order) combinations

$$
\mathcal{R} = \frac{A}{2} + H\delta u , \quad \zeta = \frac{\dot{A}}{2} - H\frac{\delta \rho}{\rho}
$$

\[^5\]This lucky fact is because the higher order terms in $N$ and $N^i$ will be multiplying the constraint equations. In particular: the third order term of $N$ and $N^i$ multiplies the zeroth order constraint equations, and the second order the first order constraint equations [13]. If we were, however, to try and generate the fourth order terms we would need $N$ and $N^i$ to second order.
CHAPTER 4. AN INFLATION MODEL FROM EFT CONSTRUCTION: SOLID INFLATION

where we have followed the notation of [43]. During our solid inflation phase, in spatially flat slicing gauge these are given by

$$\mathcal{R} = -\frac{k}{3H\epsilon} \frac{\dot{\pi}_L + H\epsilon \pi_L}{1 + k^2/3a^2H^2\epsilon}, \quad \zeta = \frac{1}{3} \nabla \cdot \pi.$$  (4.57)

where the non-local piece of $\mathcal{R}$ comes from solving the constraint equation for $N_L$.

Two peculiarities concerning the behavior of these variables during solid inflation are worth mentioning at this point. First, $\mathcal{R}$ and $\zeta$ do not coincide on super-horizon scales. Second, neither of them is conserved. These properties are in sharp contrast with what happens for adiabatic perturbations in standard cosmological models, and stem from the fact that during solid inflation, there are no adiabatic modes of fluctuation! We will clarify why this is the case in Section 4.7.

4.4 Two-point functions

Upon plugging the expressions (4.49)–(4.51) back into the action, the quadratic action for tensor, vector, and scalar fluctuations reads:

$$S^{(2)} = S^{(2)}_\gamma + S^{(2)}_T + S^{(2)}_L$$  (4.58)

$$S^{(2)}_\gamma = \frac{1}{4} M_{Pl}^2 \int dt d^3 x \, a^3 \left[ \frac{k^2/4}{1 - k^2/4a^2H} \left| \dot{\gamma}_{ij} \right|^2 + \frac{\dot{H}_{c_T}^2 k^2}{H} \left| \Pi_{ij} \right|^2 \right]$$  (4.59)

$$S^{(2)}_T = M_{Pl}^2 \int dt \int \frac{d^3 k}{(2\pi)^3} a^3 \left[ \frac{k^2/3}{1 - k^2/3a^2H} \left| \Pi_L - \left( \frac{\dot{H}}{H} \right) \Pi_L \right|^2 + \frac{\dot{H}_{c_L}^2 k^2}{H} \left| \Pi_L \right|^2 \right]$$  (4.60)

$$S^{(2)}_L = M_{Pl}^2 \int dt \int \frac{d^3 k}{(2\pi)^3} a^3 \left[ \frac{k^2/4}{1 - k^2/4a^2H} \left| \Pi_T \right|^2 + \frac{\dot{H}_{c_T}^2 k^2}{H} \left| \Pi_T \right|^2 \right]$$  (4.61)

The general (i.e. before gauge fixing) perturbed metric (to the linear-order) is parametrized by

$$g_{ij} = a(t)^2 (\delta_{ij}(1 + A) + \partial_i \partial_j \chi + \partial_i C_j + \partial_j C_i + D_{ij})$$  (4.53)

with $\partial_i C_i = 0$ and $\partial_i D_{ij} = D_{ii} = 0$; furthermore the energy momentum tensor is decomposed into scalar, vector, and tensor modes as

$$\delta T_{00} = -\bar{\rho} \delta g_{00} + \delta \rho$$  (4.54)

$$\delta T_{0i} = \bar{\rho} \delta g_{i0} - (\bar{\rho} + \bar{p})(\partial_i \delta u + \delta u_i^V)$$  (4.55)

$$\delta T_{ij} = \bar{p} \delta g_{ij} + a^2 (\delta_{ij} \delta p + \partial_i \partial_j \delta \sigma + \partial_i \delta \sigma^i + \partial_j \delta \sigma^j + \delta \sigma^i_j).$$  (4.56)
Notice the quite nontrivial $k$-dependence for $S_T^{(2)}$ and $S_L^{(2)}$ in Fourier space, which would translate into a (spatially) non-local structure in real space.

### 4.4.1 Tensor perturbations

Using (4.59) we can calculate the two-point function of the tensor perturbations. As usual, it is a simpler calculation than the scalar case and will serve as a warmup. We decompose the tensor modes into their polarizations

$$\gamma_{ij}(\vec{k}, t) = \sum_{s=\pm} \epsilon^s_{ij}(\vec{k}) \gamma^s(\vec{k}, t), \quad (4.62)$$

with $\epsilon^s_{ij} \epsilon^{s'*}_{ij} = 2 \delta^{ss'}$. The transverse, traceless conditions on $\gamma_{ij}$ now simply become $\epsilon_{ii} = k_i \epsilon_{ij} = 0$. We further decompose each $\gamma^s(\vec{k}, t)$ as

$$\gamma^s(\vec{k}, t) = \gamma^s_{cl}(\vec{k}, t) \ a^s(\vec{k}) + \gamma_{cl}^s(\vec{k}, t)^* \ a^{s\dagger}(\vec{k}^\prime). \quad (4.63)$$

where $a^s(\vec{k})^\dagger$ and $a^s(\vec{k})$ are creation and annihilation operators obeying the usual commutation relation

$$[a^s(\vec{k}), a^{s\dagger}(\vec{k}^\prime)] = (2\pi)^3 \delta^3(\vec{k} - \vec{k}^\prime) \delta^{ss'}, \quad (4.64)$$

and where the classical solution $\gamma^s_{cl}(\vec{k}, t)$ obeys the equations of motion obtained by varying (4.59):

$$d^2 d^2 \gamma_{cl} + 2aH \frac{d}{d\tau} \gamma_{cl} + (k^2 + 4\epsilon c^2 T \tau^2) \gamma_{cl} = 0. \quad (4.65)$$

In the above we have used conformal time $\tau$, where $d\tau = dt/a$, and the definition of the first slow-roll parameter (4.17). Using the time-dependence of $aH$, $\epsilon$, and $c_T$—which is worked out in Appendix G—we can express the e.o.m. for the tensor mode (4.65) up to first order in slow-roll parameters as

$$-\frac{d^2}{d\tau^2} \gamma_{cl} - \frac{2 + 2\epsilon c}{\tau} \frac{d}{d\tau} \gamma_{cl} + \left( k^2 + \frac{4\epsilon c^2 T \tau^2}{\tau^2} \right) \gamma_{cl} = 0. \quad (4.66)$$

The subscript “c” denotes that the parameters $c_T$, $\epsilon$ are evaluated at some reference time $\tau_c$, which is chosen to be the (conformal) time when the longest mode of observational relevance today exits the horizon, i.e. $\tau_c$ is defined such that

$$|c_{L,c} k_{\text{min}} \tau_c| \simeq |c_{L,c} \tau_c H_{\text{today}}| = 1 \quad (4.67)$$
where the usual convention for flat spacetime $a_{\text{today}} = 1$ is understood.

The most general solution to the above equation takes the form

$$
\gamma_{d}(\vec{k}, \tau) = (-\tau)^{3/2 + \epsilon_{c}} [A H_{\nu_{T}}^{(1)}(-k\tau) + B H_{\nu_{T}}^{(2)}(-k\tau)] , \quad \nu_{T} \simeq \frac{3}{2} + \epsilon_{c} - \frac{4}{3} c_{T}^{2} \epsilon_{c} , \quad (4.68)
$$

where $H^{(1,2)}$ are Hankel functions, and $A$ and $B$ are constants to be fixed by matching the appropriate initial conditions.

At very early times the physical wavelength $-k/a$ is so small compared to the Hubble scale $H$ that the curvature of spacetime cannot be perceived by such modes; it is therefore expected that the canonically normalized classical solution should match the free wave function in the flat-space vacuum, $\frac{1}{\sqrt{2}k} e^{-i k \tau}$. Note that the canonically normalized field $\gamma_{\text{can}}^{s}(\vec{k}, \tau)$ is related to $\gamma^{s}(\vec{k}, \tau)$ by

$$
\gamma^{s}(\vec{k}, \tau) = \frac{\sqrt{2}}{M_{\text{Pl}}} \gamma_{\text{can}}^{s}(\vec{k}, \tau) . \quad (4.69)
$$

Thus, the initial condition for $\gamma_{d}^{s}(\vec{k}, \tau)$ is specified by

$$
\lim_{\tau \to -\infty} \gamma_{d}^{s}(\vec{k}, \tau) = \frac{1}{\sqrt{2}k M_{\text{Pl}} a(\tau)} e^{-i k \tau} . \quad (4.70)
$$

Comparing this to the general solution given by (4.68) and using the asymptotic form (for large arguments) of Hankel functions,

$$
\lim_{x \to +\infty} H_{m}^{(1)}(x) \to \sqrt{\frac{2}{\pi x}} e^{-ixi\left(m + \frac{1}{2}\right)\frac{\pi}{2}}, \quad \lim_{x \to +\infty} H_{m}^{(2)}(x) \to \sqrt{\frac{2}{\pi x}} e^{-ix + i\left(m + \frac{1}{2}\right)\frac{\pi}{2}} . \quad (4.71)
$$

we fixed

$$
A = \frac{H_{c}(1 - \epsilon_{c})}{M_{\text{Pl}}} \sqrt{\frac{\pi}{2}} \left(-\tau_{c}\right)^{-\epsilon_{c} e^{i(\nu_{T} \pi/2 + \pi/4)}} + \mathcal{O}(\epsilon^{2}) , \quad B = 0 . \quad (4.72)
$$

where, once again, $H_{c} \equiv H(\tau_{c})$ and $\epsilon_{c} = \epsilon(\tau_{c})$. This result is valid up to first order in slow roll.

It is interesting to note that even when these tensor modes are well outside the horizon, they are not conserved. A similar story applies to the gauge invariant curvature perturbations $\zeta$ and $\mathcal{R}$ defined by (4.57) and the vector perturbation (like $\pi_{T}^{i}$), as we will see in the following section. This is in contrast to the usual situation in most inflation models, and will be discussed in more detail in Section 4.7. In particular, by utilizing the asymptotic limit (for small argument) of the Hankel function

$$
\lim_{x \to 0^{+}} H_{m}^{(1)}(x) \to (-i)^{m} \frac{\Gamma(m)}{\pi} \left(\frac{2}{x}\right)^{m} , \quad (4.73)
$$
the mild time-dependence of the tensor mode at late time is given by:

$$\lim_{-k\tau \to 0^+} \gamma^s_{cl}(\vec{k}, \tau) = k^{-3/2} \left( \frac{\tau}{\tau_c} \right)^{4c^2_{L,e}c^{s_1s_2}/3} (-k\tau_c)^{c^2_{L,e}c^{s_1s_2}} \left( \frac{iH_e}{M_{Pl}} + O(\epsilon) \right) .$$  \hspace{1cm} (4.74)$$

where we have made use of relation (4.27). As we will see soon, the transverse vector modes and scalar modes in our model share this feature as well.

And so, finally, we are ready to obtain the two-point function for the tensor perturbations of the metric. In particular, we are interested in its late time behavior, when modes are well outside the horizon:

$$\langle \gamma_{s_1}(\vec{k}_1, \tau) \gamma_{s_2}(\vec{k}_2, \tau) \rangle = \left( 2\pi \right)^3 \delta^3(\vec{k}_1 + \vec{k}_2) \delta^{s_1s_2} |\gamma_{cl}(\vec{k}, \tau)|^2$$  \hspace{1cm} (4.75)$$

The dependence on \( k \) and \( \tau \) is kept to first order in slow roll while the overall constant is to lowest order.

The advantage of expressing time-dependent quantities in reference to a fixed fiducial time \( \tau_c \), as opposed to the usual convention of using the time at horizon crossing, is that the time- and momentum-dependence are made manifest. We can simply read off the tilt of the spectrum to first order in slow roll from the above expression:

$$n_T - 1 \simeq 2c^2_{L,e}c^{s_1s_2} .$$  \hspace{1cm} (4.76)$$

We can see that the two point function for tensor modes is blue shifted, which matches the result of [59], and which is a distinctive signature of our scenario, unreproducible by more conventional models of inflation. As to the spectrum’s overall amplitude, it is the usual one: \( \langle \gamma\gamma \rangle \sim H^2/M_{Pl}^2 \).

### 4.4.2 Scalar Perturbations

We proceed by calculating the scalar two point function in a similar manner as above. As emphasized in section 4.3, the scalar quantity of interest\(^7\) is the gauge invariant quantity \( \zeta \),

\(^7\)We find \( \zeta \) a more interesting quantity than \( R \) for the reason that given our assumptions about reheating, \( \zeta \) evolves continuously from inflation to the post-inflationary phase, while \( R \) does not. See Section 4.8 for details. However, \( R \) does play a vital role as a simplifier in solving for the classical solution for scalar perturbation.
which, in Fourier space, is related to the longitudinal Goldstones $\pi_L$ simply by $\zeta = -k\pi_L/3$ (see eq. (4.57)).

As before, let’s decompose the scalar field of interest in terms of creation and annihilation operators:

$$\zeta(\vec{k}, t) = \zeta_{cl}(\vec{k}, t) b(\vec{k}) + \zeta_{cl}(\vec{k}, t)^* b^\dagger(-\vec{k}) ,$$

where the usual commutation relation is obeyed $[b(\vec{k}), b^\dagger(\vec{k}')] = (2\pi)^3\delta^3(\vec{k} - \vec{k}')$.

The classical equation of motion for $\zeta_{cl}$ follows from varying the quadratic $\pi_L$ action (4.61). The general solution to this equation is quite complicated, however there is a trick that makes its solution much easier. If we re-express the e.o.m. in terms of the other gauge invariant parameter $R_{cl}$ (see eq. (4.57)) we have

$$-3c_L^2 \zeta_{cl}(\vec{k}, t) = \frac{1}{H} \dot{R}_{cl}(\vec{k}, t) + (3 + \eta(t) - 2\epsilon(t)) R_{cl}(\vec{k}, t) .$$

which, together with the definition of $R$

$$R_{cl} = \frac{1}{H\epsilon} \frac{\dot{\zeta}_{cl} + H\epsilon \zeta_{cl}}{1 + k^2/3a^2H^2\epsilon} ,$$

forms a system of two first order equations of two variables. Eliminating $\zeta_{cl}$ we can generate a second order equation for $R_{cl}$ which takes the usual form similar to (4.65). Written with respect to conformal time, and up to first order in slow roll, it is given by:

$$R_{cl}'' + (2 + \eta - 2s) aH R_{cl}' + [c_L^2 k^2 + (3\epsilon - 6s + 3c_L^2\epsilon)a^2H^2]\ R_{cl} = 0 ,$$

where prime denotes a derivative w.r.t. conformal time, and $s$ is the slow roll parameter defined by (4.40). Once again, using the conformal time dependence of $aH$, $s$, $\eta$, $\epsilon$, and $c_L$ contained in Appendix G this equation can be solved in terms of Hankel functions. One finds that the general solution to first order in the slow roll parameters is given by

$$R_{cl}(\vec{k}, \tau) = (-\tau)^{-\alpha} \left[ \mathcal{C} H^{(1)}_{\nu_S}(-c_L(\tau)k\tau(1 + s_c)) + \mathcal{D} H^{(2)}_{\nu_S}(-c_L(\tau)k\tau(1 + s_c)) \right]$$

where $\alpha = -\frac{1}{2}(3 + 2\epsilon_c + \eta_c - 2s_c)$ and $\nu_S = \frac{1}{2}(3 + 5s_c - 2c_L^2\epsilon_c + \eta_c)$. Notice that for this to be a solution, it is important to keep into account—to first order—the time-dependence of $c_L$ in the argument of the Hankel functions.
Once again, in order to match the initial conditions we must canonically normalize $\pi_L$.

A quick glance at (4.61) reveals that the correct canonically normalized field is

$$\pi^\text{can}_L(\vec{k},\tau) = \sqrt{2} \pi_L \left[ \frac{M^2_{\text{Pl}} a^2 k^2}{3 \left( 1 + \frac{k^2}{3a^2 H^2} \right)} \right]^{1/2} \xrightarrow{\kappa \to \infty} \sqrt{2} e M_{\text{Pl}} a^2 \pi_L.$$  \hspace{1cm} (4.82)

With the usual normalization for the creation and annihilation operators we will recover the Minkowski vacuum for very early times by demanding that

$$\lim_{\tau \to -\infty} \zeta_{cl}(\vec{k},\tau) = -\frac{k \pi^\text{can}_{cl}}{3\sqrt{2} e M_{\text{Pl}} a^2} = -\sqrt{\frac{k}{4 e c_L}} \frac{e^{-i(1 + s_c)c_L(\tau)k\tau}}{3 M_{\text{Pl}} a^2}. \hspace{1cm} (4.83)$$

Or, equivalently that

$$\lim_{\tau \to -\infty} \mathcal{R}_{cl}(\vec{k},\tau) = \frac{a^2 H^2}{k} \frac{d}{d\tau} \left( \frac{\pi^\text{can}_{cl}}{H} \right) = i \sqrt{\frac{c_L}{4 e k}} \frac{e^{-i(1 + s_c)c_L(\tau)k\tau}}{M_{\text{Pl}} a}. \hspace{1cm} (4.84)$$

Matching the general solution given by (4.81) to the the initial condition (4.84) will set $\mathcal{D} = 0$ and

$$\mathcal{C} = -i \sqrt{\frac{\pi}{8 e c}} \frac{c_{L,c} H_c}{M_{\text{Pl}}} (\tau_c) e^{-\epsilon_c - \eta_c/2} (1 + \frac{1}{2} s_c - \epsilon_c) e^{i(-\eta_c + 5s_c - 2c_{L,c}^2)\pi/4} + \mathcal{O}(e^{3/2}). \hspace{1cm} (4.85)$$

One can now use (4.78) to obtain the full expression for $\zeta_{cl}(\vec{k},\tau)$, which is (as promised) a bit messy and not particularly instructive as for our computation we are only interested in $\zeta$'s late time limit. We will not bother to write it out here.

Just like the tensor perturbations, neither $\mathcal{R}$ nor $\zeta$ is conserved outside the horizon, though their temporal dependence is mild, i.e., suppressed by slow-roll parameters:

$$\lim_{{-k\tau \to 0^+}} \mathcal{R}_{cl}(\vec{k},\tau) = \left( \frac{\tau}{\tau_c} \right)^{\frac{4}{3} c^2_{L,c} \epsilon_c - 2s_c} (-c_{L,c} k\tau_c) c^2_{L,c} \epsilon_c - 5s_c/2 - \eta_c/2 \left( -\frac{H_c}{\sqrt{4 e c M_{\text{Pl}} c_{L,c}^2 k^{3/2}}} + \mathcal{O}(e^{1/2}) \right) \hspace{1cm} (4.86)$$

$$\lim_{{-k\tau \to 0^+}} \zeta_{cl}(\vec{k},\tau) = \left( \frac{\tau}{\tau_c} \right)^{\frac{4}{3} c^2_{L,c} \epsilon_c} (-c_{L,c} k\tau_c) c^2_{L,c} \epsilon_c - 5s_c/2 - \eta_c/2 \left( \frac{H_c}{\sqrt{4 e c M_{\text{Pl}} c_{L,c}^5/2 k^{3/2}}} + \mathcal{O}(e^{1/2}) \right). \hspace{1cm} (4.87)$$

Notice that at this order in slow-roll, on large scales $\zeta$ and $\mathcal{R}$ are proportional to each other, with proportionality constant $c^2_{L,c}$, which is in agreement with (4.78):

$$\mathcal{R} \simeq -c^2_{L,c}(\tau) \zeta \quad (k\tau \to 0^-). \hspace{1cm} (4.88)$$
Now finally, the two point function of $\zeta$ for late times (when the modes are well outside the horizon) is given by

$$ \langle \zeta(\tau, \vec{k}_1)\zeta(\tau, \vec{k}_2) \rangle = (2\pi)^3 \delta^3(\vec{k}_1 + \vec{k}_2) \left| \zeta_{cd}(\tau, \vec{k}_1) \right|^2 $$

(4.89)

where, as before, we have kept the dependence on $k$ and $\tau$ to first order in slow roll while the prefactor is expressed only to lowest order in slow roll.

Once again, since all the parameters are evaluated at the global time $\tau_c$ as opposed to the time of horizon crossing for each mode, we can simply read off the tilt to first order in slow roll directly from the above expression. It is:

$$ n_S - 1 \simeq 2 \epsilon_c c_{L,c}^2 - 5 s_c - \eta_c $$

(4.91)

Notice the overall $1/c_L^5$ factor in front of the spectrum. In a more standard single-field model, this would be replaced by $1/c_L$ (see e.g. [30]). For small $c_L$, our extra powers of $c_L$ give us a very suppressed tensor-to-scalar ratio:

$$ r \sim \epsilon c_L^5 $$

(4.92)

It is crucial however to ascertain whether we should focus on the $\zeta\zeta$ spectrum or the $\mathcal{R}\mathcal{R}$ one; they differ by a factor of $c_L^4$ as can be seen from (4.88). In Section 4.8 we will argue that it is more appropriate to focus on the $\zeta\zeta$ spectrum, given our model for reheating.

### 4.5 The three-point function

We now compute the $\langle \zeta\zeta\zeta \rangle$ three-point function. Like in single-field models with a small speed of sound, our three-point function will be enhanced by inverse powers of $c_L$ with respect to what one gets in standard slow-roll inflation, for essentially the same reason (see e.g. [30]). However in addition to this, we will find an extra $1/\epsilon$ enhancement, coming from the fact that the quadratic phonon action (4.21) is suppressed by $\epsilon$, whereas the cubic interactions are not.

In order to compute the correlation function at a specific time, we need to evolve it from a quantum state we know, that is the early-time flat-space vacuum. Expanding the
usual time-evolution operator and working to lowest order in perturbation theory we have
the standard result, which is given schematically by:
\[ \langle \zeta(\tau)^3 \rangle = -i \int_{-\infty}^{\tau} d\tau' \langle \Omega(-\infty) | \langle \zeta(\tau)^3, H_{\text{int}}(\tau') \rangle | \Omega(-\infty) \rangle . \] (4.93)

For our purposes it is enough to calculate the three-point function to lowest order in slow-
roll. As demonstrated in Appendix \[I\] this order it is enough to work with the phonon cubic
action, which in our FRW curved background takes the form (neglecting boundary
terms):
\[ \mathcal{L}_3 = M_{\text{Pl}}^2 a(t)^3 H^2 \frac{F_Y}{F} \left\{ \frac{7}{81} (\partial \pi)^3 - \frac{1}{5} \partial \pi \partial_j \pi^k \partial_k \pi^j - \frac{4}{9} \partial \pi \partial_j \pi^k \partial_j \pi^k + \frac{2}{3} \partial \pi \partial_j \partial_j \partial_k \partial_k \pi^j \right\} . \] (4.94)

Quite amazingly, this applies both in the decoupling limit \( k \gg a H_{c}^{1/2} \) and in the opposite
limit \( k \ll a H_{c}^{1/2} \). And so, defining \( \zeta_i \equiv \zeta(\tau, \vec{k}_i) \), we have
\[ \langle \zeta_1 \zeta_2 \zeta_3 \rangle = i M_{\text{Pl}}^2 F_Y k_1 k_2 k_3 \frac{27}{27} \int_{\vec{p}_1, \vec{p}_2, \vec{p}_3} (2\pi)^3 \delta^3(\vec{p}_1 + \vec{p}_2 + \vec{p}_3) Q(\vec{p}_1, \vec{p}_2, \vec{p}_3) \times \]
\[ \int_{-\infty}^{\tau} d\tau' a'(\tau') H^2(\tau') \left[ \pi_L(\tau', \vec{k}_1) \pi_L(\tau', \vec{k}_2) \pi_L(\tau', \vec{k}_3), \pi_L(\tau, \vec{p}_1) \pi_L(\tau, \vec{p}_2) \pi_L(\tau, \vec{p}_3) \right] , \] (4.95)
where
\[ Q(\vec{p}_1, \vec{p}_2, \vec{p}_3) \equiv \frac{7}{81} p_1 p_2 p_3 - \frac{5}{27} \left( \frac{p_1 (\vec{p}_2 \cdot \vec{p}_3)^2}{p_2 p_3} + \frac{p_2 (\vec{p}_1 \cdot \vec{p}_3)^2}{p_1 p_3} + \frac{p_3 (\vec{p}_1 \cdot \vec{p}_2)^2}{p_1 p_2} \right) + \frac{2}{3} \frac{(\vec{p}_1 \cdot \vec{p}_2)(\vec{p}_2 \cdot \vec{p}_3)(\vec{p}_3 \cdot \vec{p}_1)}{p_1 p_2 p_3} , \] (4.96)
which is totally symmetric under permutations of \( \vec{p}_1, \vec{p}_2, \vec{p}_3 \).

Writing \( \pi_L \) in terms of creation and annihilation operators allows us to easily express
the integral in terms of the classical solutions. To be precise,
\[ \langle \zeta_1 \zeta_2 \zeta_3 \rangle = -(2\pi)^3 \delta^3(\vec{k}_1 + \vec{k}_2 + \vec{k}_3) \times 6 M_{\text{Pl}}^2 F_Y k_1 k_2 k_3 \frac{27}{27} Q(\vec{k}_1, \vec{k}_2, \vec{k}_3) I(\tau; -\infty) , \] (4.97)
where the integral \( I(\tau_1; \tau_2) \) is defined as
\[ I(\tau_1; \tau_2) = J(\tau_1; \tau_2) + J^*(\tau_1; \tau_2) \] (4.98)
\[ J(\tau_1; \tau_2) \equiv -i \pi^c_L(\tau_1, \vec{k}_1) \pi^c_L(\tau_1, \vec{k}_2) \pi^c_L(\tau_1, \vec{k}_3) \int_{\tau_2}^{\tau_1} d\tau' \left( \frac{\tau'/\tau_c - 2e}{H^2_2 \tau'^4} \right) \pi^c_L(\tau', \vec{k}_1) \pi^c_L(\tau', \vec{k}_2) \pi^c_L(\tau', \vec{k}_3) \]
and we used that \( Q \) is an even function of the momenta. Just as in the previous section,
we have used the dependence of \( H \) and \( a \) on conformal time with reference to \( \tau_c \) to lowest
order in slow roll.
With the form of the full classical solution $\pi^cl_L$ — which can be straightforwardly derived from the classical solution of $R(\tau, \vec{k})$, Eqn. (4.81) — we are not able to perform the integral (4.98) analytically. So we need to look for some computationally simple perturbative expression for $\pi^cl_L$. Notice that at sufficiently early times, $c_L k \tau \gtrsim O(\epsilon)$, thus slow-roll parameters — e.g. $\epsilon, \eta, s$ which are all assumed to be of order $O(\epsilon)$ — serve as good expansion parameters and we only need to retain $O(\epsilon^0)$ terms. While at late times $c_L k \tau \lesssim O(\epsilon)$, the role as the expansion parameter is replaced by $c_L k \tau$, and we keep in the expression for $\pi$ only the leading order term in the $c_L k \tau$ expansion. It turns out that if we are only interested in the three point function to lowest order in slow rolls, we can just use the two asymptotic profiles of $\pi^cl_L$ and ‘glue’ them together smoothly in the intermediate regime:

$$\pi^cl_L(\tau, \vec{k}) \simeq \begin{cases} B_k (1 + ic_L k \tau - \frac{1}{3}c_L^2 k^2 \tau^2) e^{-ic_L k \tau}, & \text{for } |c_L k \tau| \gtrsim \epsilon \, (4.99a) \\ B_k (-c_L,c k \tau)^{c_L,c,\epsilon+c} (c_L,c k \tau e^{-5s_\epsilon/2-\eta_\epsilon/2-\epsilon}), & \text{for } |c_L k \tau| \lesssim \epsilon \, (4.99b) \end{cases}$$

where

$$B_k = -\frac{3}{2} \frac{H_c}{M_{Pl}^{5/2} c^{1/2} k^{5/2}}. \quad (4.100)$$

Our strategy now is to break up the integral (4.98) into separate regions where one of the functional forms described by (4.99a) and (4.99b) can used. The integral can then be done explicitly.

### 4.5.1 Analytic Calculation of Integral

To illustrate the point and make the flavor of the analysis transparent, let’s look at an approximately equilateral configuration of momenta. That is, assume that

$$\vec{k}_1 + \vec{k}_2 + \vec{k}_3 = 0, \quad k_1 \sim k_2 \sim k_3 \sim k. \quad (4.101)$$

First, notice that given some reference time $\tau_*$, we can split the time-integral as

$$J(\tau; -\infty) = \frac{\pi^cl_L(\tau, \vec{k}_1)\pi^cl_L(\tau, \vec{k}_2)\pi^cl_L(\tau, \vec{k}_3)}{\pi^cl_L(\tau_*, \vec{k}_1)\pi^cl_L(\tau_*, \vec{k}_2)\pi^cl_L(\tau_*, \vec{k}_3)} J(\tau_*; -\infty)$$

$$- i \pi^cl_L(\tau, \vec{k}_1)\pi^cl_L(\tau, \vec{k}_2)\pi^cl_L(\tau, \vec{k}_3) \int_{\tau_*}^{\tau} d\tau' \frac{(\tau'/\epsilon)^{-2s_\epsilon}}{H_c^2 \epsilon^4} \pi^cl_*(\tau', \vec{k}_1)\pi^cl_*(\tau', \vec{k}_2)\pi^cl_*(\tau', \vec{k}_3). \quad (4.102)$$

Then, choosing $\tau_*$ to be precisely the conformal time at which a mode of momentum $k$ transitions from (4.99a) to (4.99b), $-c_L k \tau_* \sim \epsilon$, we find that the real part of the second
line vanishes at zeroth order in $\epsilon$, because all the $\pi_L$’s involved in the expression—inside and outside the integral—are real, and there is an overall $i$. The remaining piece is all that will contribute to the integral. And so we can write

$$J(\tau; -\infty) + J^*(\tau; -\infty) = \prod_{i=1}^{3} |B_{k_i}|^2 \left( -c_L k_i \tau \right)^{\frac{3}{2} \epsilon + \epsilon} \left( -c_L k_i \tau - 5s/2 - \eta/2 - \epsilon \right) \times$$

$$\int_{-\infty}^{\tau} d\tau' \frac{-i(\tau'/\tau_c)^{\epsilon}}{H_c^2 \tau'^4} \prod_{j=1}^{3} \left( 1 - ic_L k_j \tau' - \frac{1}{4} c_L^2 k_j^2 \tau'^2 \right) e^{+ic_L k_j \tau'} + c.c.$$

$$= - \frac{1}{27} \frac{c_L^3}{H_c^3} k_1 k_2 k_3 U(k_1, k_2, k_3) \left( \frac{\tau}{\tau_c} \right)^{4c_L^2 \epsilon} \prod_{i=1}^{3} |B_{k_i}|^2 (-c_L k_i \tau_c)^{\frac{3}{2} \epsilon - 5s/2 - \eta/2},$$

(4.103)

where the scale invariant function $U(k_1, k_2, k_3)$ is given by

$$U(k_1, k_2, k_3) = \frac{2}{k_1 k_2 k_3 (k_1 + k_2 + k_3)^3} \left\{ 3(k_1^6 + k_2^6 + k_3^6) + 20 k_1^2 k_2^2 k_3^2 \right.$$  

$$+ 18(k_1^4 k_2^2 k_3 + k_1^2 k_2^4 k_3 + k_1 k_2^2 k_3^4) + 12(k_1^3 k_2^3 + k_3^2 k_3^3 + k_3^3 k_1^3)$$  

$$+ 9(k_1^5 k_2 + 5 \text{ perms}) + 12(k_1^4 k_2^2 + 5 \text{ perms}) + 18(k_1^3 k_2^3 k_3 + 5 \text{ perms}) \right\}.$$

(4.104)

In order to ensure convergence of the integral and project onto the right vacuum, the integral is actually computed over a slightly tilted contour, that is $\tau' \rightarrow (1 - i\epsilon) \tau' + \tau_s$, with $\epsilon \rightarrow 0^+$, and the limits of integration are from $-\infty$ to 0. Additionally, in the last step, the fact that $1 \gg |c_L k \tau_s| \sim \epsilon$, and $|\tau_s| > |\tau|$ has been used to collect only the leading order in slow roll contributions.

A more careful analysis of the same flavor applies to more general triangle shapes, see [25]. It turns out that the above expression is valid provided that the triangle formed by the various momenta is not too squeezed, that is, provided

$$k_{\text{long}}/k_{\text{short}} > \sqrt{\epsilon}.$$  

(4.105)

And so, finally, putting everything together, we can express the full three-point function as

$$\langle \zeta_1 \zeta_2 \zeta_3 \rangle (\tau_c) \simeq (2\pi)^3 \delta^3(\vec{k}_1 + \vec{k}_2 + \vec{k}_3) \times$$

$$\frac{3 F_Y}{32 F} H_c^4 \frac{M_{Pl}}{e^3 c_L^3} \frac{1}{\tau_c} \left( \frac{\tau_c}{\tau_c} \right)^{4c_L^2 \epsilon} \frac{Q(\vec{k}_1, \vec{k}_2, \vec{k}_3) U(k_1, k_2, k_3)}{k_1^4 k_2^2 k_3^3},$$

(4.106)
where, we remind the reader, $Q(k_1, k_2, k_3)$ is given by (4.96) and $U(k_1, k_2, k_3)$ is given by (4.104).

We will remark that the mild time dependence, $(\tau_e/\tau_c)^{4/2} \epsilon$, in the above expression can actually produce an order one correction to the overall magnitude of the three-point function. Indeed, assuming that inflation lasts for $N_e \sim 60$ e-folds after the longest mode of today’s relevance exits the horizon, we can see immediately that $(\tau_e/\tau_c)^{O(\epsilon)} \sim e^{-60 \times O(\epsilon)}$, which, as promised, depending on how small $\epsilon$ is, can give an $O(1)$ correction. On the other hand, the mild momentum dependence $(-c_L k_i \tau_c)^{c_2^2} e^{-5a/2-\eta/2}$, appearing in (4.103), is equal to one up to $O(\epsilon)$ corrections. We thus drop this piece from (4.106), in order to be consistent with the preceding computation of the integral. Our result (4.106) should be understood as the leading order contribution in slow roll.

4.6 The size and shape of non-gaussianities

It is useful to rewrite the three-point function above as an overall amplitude $f_{NL}$ times a shape that is a function of the momenta with order-one coefficients [61]. It is customary to do so at the level of correlators of the Newtonian potential $\Phi$ during matter-domination, rather than of $\zeta$. The relation outside the horizon is

$$\Phi = \frac{3}{5} \zeta \; .$$

Neglecting the tilt, one then writes the two- and three-point functions as

$$\langle \Phi(k_1) \Phi(k_2) \rangle = (2\pi)^3 \delta^3(k_1 + k_2) \cdot \frac{\Delta \Phi}{k_1^3}$$

$$\langle \Phi(k_1) \Phi(k_2) \Phi(k_3) \rangle = (2\pi)^3 \delta^3(k_1 + k_2 + k_3) \cdot f(k_1, k_2, k_3) \; ,$$

and normalizes $f$ on ‘equilateral’ configurations with $k_1 = k_2 = k_3$ [63],

$$f(k, k, k) = f_{NL} \cdot \frac{6 \Delta \Phi^2}{k^6} \; .$$

This defines the parameter $f_{NL}$ in a model-independent fashion, in terms of observable quantities only. In particular, the observed value for the power-spectrum normalization is $\Delta \Phi \simeq 2 \times 10^{-8}$. Notice that, because of momentum conservation, the three momenta $k_{1,2,3}$
close into a triangle. As a result, the function $f$ depends on the absolute values $k_{1,2,3}$ only, because a triangle is uniquely defined—up to overall rotations, which are a symmetry of $f$—by specifying its sides. Notice also that scale-invariance forces $f$ to have overall scaling dimension $k^{-6}$, and we have used this fact in (4.110). Notice finally that the standard convention would be to call $F$ the function that we call $f$. Unfortunately we have already been using $F$ for our Lagrangian, so will stick to $f$ for the function defined above. Hopefully this will not lead to confusion.

Applying these definitions to our case we find

$$
\Delta \Phi = \frac{9}{100} \left( \frac{\tau_e}{\tau_c} \right)^{8c_2^2/3} \cdot \frac{H^2}{M_{Pl}^2} \cdot \frac{1}{\epsilon c_L^2} 
$$

(4.111)

$$
f(k_1, k_2, k_3) = \frac{5}{2} \cdot \frac{F_Y}{F} \cdot \frac{1}{\epsilon c_L^2} \cdot \left( \frac{\tau_e}{\tau_c} \right)^{-4c_2^2/3} \cdot \Delta_{\Phi}^2 \cdot \frac{Q(k_1, k_2, k_3) \cdot U(k_1, k_2, k_3)}{k_1^3 k_2^3 k_3^3} 
$$

(4.112)

$$
f_{NL} = -\frac{19415}{13122} \cdot \left( \frac{\tau_e}{\tau_c} \right)^{-4c_2^2/3} \cdot \frac{F_Y}{F} \cdot \frac{1}{\epsilon c_L^2} \simeq -O(1) \cdot \frac{F_Y}{F} \cdot \frac{1}{\epsilon c_L^2} 
$$

(4.113)

The $f_{NL}$ parameter gives us a measure of the absolute size of non-gaussianities. As we argued in Section 4.2, $F_Y$ is essentially a free parameter, which can be as large as $F$, in which case our $f_{NL}$ is huge, of order $1/(\epsilon c_L^2)$. By comparison, single-field inflationary models with small sound speed—whose non-gaussianities are much larger than for standard slow-roll inflation—have, at the same value of the sound speed, an $f_{NL}$ which is a factor of $\epsilon$ smaller than ours. Notice that in this case there is a potential tension for our model: the same combination $\epsilon \cdot c_L^2$ appears in the scalar tilt, eq. (4.91). Of course one could have cancellations there, because of the other terms in the expression for the tilt. But assuming that these do not change the overall order of magnitude of the tilt, the tilt is small if non-gaussianities are large, and vice versa. Eventually, one should observe either. If on the other hand our $F_Y$ is of order $\epsilon F$, then this $1/\epsilon$ enhancement for non-gaussianities is gone, and our $f_{NL}$ becomes of order $1/c_L^2$, which is the same as for small sound-speed single-field models.

---

8We remind the reader that cosmological perturbations can still be nearly gaussian, even for huge values of $f_{NL}$, as long as the combination $f_{NL} \sqrt{\langle \zeta^2 \rangle}$ is much smaller than one at the relevant scales. For us, in the most strongly coupled case ($F_Y \sim F$), such a combination is of order $H/M_{Pl} \cdot (\epsilon c_L^2)^{-3/2}$, which is much smaller than one if and only if the weak-coupling condition (4.39) is obeyed. As usual, perturbations are nearly gaussian if and only if they are weakly coupled at horizon crossing.
Figure 4.1: The shape of non-gaussianities for our model, according to the standard conventions and definitions of ref. [61].

But the most interesting feature of our non-gaussian signal is probably its shape, that is, the dependence of $f$ on the shape of the triangle made up by the momenta $\vec{k}_{1,2,3}$. We plot it in Fig. 4.1, following the standard conventions of [61]. In particular, it is clear from the plot that our three-point function is peaked on ‘squeezed’ triangles with $k_3 \ll k_{1,2}$, but its behavior for those configurations depends strongly on the angle $\theta$ between $\vec{k}_3$ and the other momenta. Quantitatively, focusing on the $\frac{QU}{k_3 k_2}$ structure in (4.112) and ignoring the prefactors from now on, we get

$$f(k_1, k_2, k_3 \to 0) \propto -\frac{40}{27} \frac{(1 - 3 \cos^2 \theta)}{k_1^3 k_3^3}.$$  (4.114)

where we used momentum-conservation to rewrite $k_2$ as $k_2 \simeq k_1 + k_3 \cos \theta$. Such an angular dependence is not there in any of the standard inflationary models we are aware of: at least for single-field models, the consistency relations force the behavior of the three-point function in the squeezed limit to be angle-independent—see e.g. [13][66]. On the other hand, in our case the standard consistency relations are maximally violated, both at the level of angular dependence—as we just mentioned—and at the level of the overall prefactor: usually the squeezed limit of the three-point function is suppressed by the scalar tilt, which
is of order $\epsilon$; here instead, there is no suppression like that, and in fact, as we argued above the overall prefactor can be as big as one over the tilt. It is easy to see why the consistency relations do not hold in our model: the standard argument of [13, 66]—that a long-wavelength background $\zeta$ can be traded in for a rescaling of spatial coordinates—does not apply to our case, because in our model there is no gauge in which the curvature perturbation $\zeta$ appears as a $\zeta \cdot \delta_{ij}$ correction to the spatial metric.

4.7 Why Is $\zeta$ Not Conserved?

We saw in Section 4.4 that already at linear level, during our solid inflation phase neither $\zeta$ nor $R$ is conserved on large scales. One might be tempted to attribute this to the presence of isocurvature modes in addition to adiabatic ones. However, in our model there is only one scalar perturbation—parameterized by $\pi_L$ in the gauge we have been using—and usually isocurvature modes are a luxury that only multi-field models can afford. To sharpen the paradox, our $\zeta$ and $R$ do not coincide on large scales—see eq. (4.88)—whereas usually they do, even for fluctuations that are not purely non-adiabatic, that is, even when they are not conserved. In fact, Weinberg proved a no-go theorem stating that all FRW cosmological models—inflationary or not—feature two adiabatic modes of fluctuation, one of which has constant and identical $\zeta$ and $R$ on large scales, while the other has $\zeta = R = 0$ [8, 43]. This theorem is manifestly violated by our model. Before showing this, let us explain in physical terms why our solid system cannot sustain adiabatic modes.

By definition, an adiabatic mode is a perturbation that for very long wavelengths becomes locally unobservable, being indistinguishable from a slight shift in time of the background solution. An ordinary fluid offers a perfect example of this. Consider a long-wavelength sound wave in a fluid, for the moment in the absence of gravity. For an observer making measurements on scales much shorter than the wavelength, and working in the local rest frame—which is slightly different from the background one—, the only observables are the density and the pressure: neglecting gradients, a fluid is isotropic in its rest frame, and its stress-energy tensor is characterized by $\rho$ and $p$ only, which are related by the equation of state. Then, the only physical effect that is measurable on scales much shorter than the
wavelength is the local compression (or dilation) the sound wave induces. When we include gravity into the picture, essentially the same considerations apply for the perturbation, but now the time-evolution of the unperturbed FRW background already probes all possible compression levels for the fluid (within some range), that is, all possible values of $\rho$ and $p$ compatible with the equation of state. As a consequence, for wavelengths much longer than the Hubble scale, within any given Hubble patch a sound wave will be indistinguishable from a time-shift of the background solution, that is, it will become physically unobservable. Different Hubble patches will then evolve as separate identical FRW universes, and it can be shown that this translates into a conservation law for $R$.

For a solid, the situation is very different, already in the absence of gravity. A longitudinal phonon—which is the only scalar fluctuation at our disposal—does not correspond to a purely compressional deformation of the medium. Even for wavelengths that are much longer than the observation scale, the anisotropic stress and the compression associated with the phonon are of the same order of magnitude. This is evident from the form of the stress-energy tensor (4.16), which expanded to first order in $\vec{\pi}$ yields schematically

$$\delta T_{ij} \sim (\vec{\nabla} \cdot \vec{\pi}) \delta_{ij} + (\partial_i \pi_j + \partial_j \pi_i), \quad (4.115)$$

with similar coefficients in front of the two tensor structures—related to suitable derivatives of $F$ w.r.t. $X, Y, Z$—whose precise values will not concern us here. For a longitudinal phonon of momentum $\vec{q}$, the anisotropic stress is proportional to $\hat{q}_i \hat{q}_j$, and is of the same order of magnitude as the change in pressure. A local observer can detect this anisotropy if he or she can detect the change in pressure. In other words, unlike a fluid, a solid with small longitudinal deformations is not locally isotropic. Once we include gravity, these scalar fluctuations will be locally distinguishable from the background solution, even for superhorizon wavelengths, since the background is isotropic.

We can now go back to the no-go theorem of ref. [8], and see which assumptions we are violating. The theorem is based on the following ingenious idea. Since an adiabatic mode is, by definition, unobservable once the wavelength is very long, at zero momentum it should reduce to a gauge transformation of the FRW background. Newtonian gauge is a complete gauge-fixing at finite momentum, but it has a residual gauge freedom at zero momentum. By exploiting this gauge freedom one can construct zero-momentum solutions
of the linearized perturbation equations. Most of these are pure-gauge, unphysical solutions. To be physical, they have to be the zero-momentum limit of finite momentum solutions, which are physical because there is no residual gauge-freedom at finite momentum. For this to be the case, one needs the zero-momentum solution to obey the finite-momentum version of the \((ij)\) and \((0i)\) linearized Einstein equations,

\[
\Phi = \Psi - (8\pi G) \delta \sigma \\
\dot{H} \delta u = \dot{\Psi} + H \Phi ,
\]

which singles out only two independent modes among the zero-momentum solutions. Here \(\delta \sigma\) and \(\delta u\) are the scalar anisotropic stress and the velocity potential of eqs. (4.55), (4.56),

\[
\delta T_{ij} \supset a^2(t) \partial_i \partial_j \delta \sigma , \quad \delta T_{0i} \supset -(\bar{\rho} + \bar{p}) \partial_i \delta u .
\]

One has to impose further that all equations of motion—for gravity and for the matter fields—are regular for \(\vec{q} \to 0\), including eq. (4.116). More precisely, if one rewrites all linearized equations of motion in first-order form,

\[
\dot{y}_a(\vec{q}, t) + C_{ab}(\vec{q}, t) \cdot y_b(\vec{q}, t) = 0 ,
\]

where the \(y_a\)'s include the fields and their velocities, with constraints for the initial conditions of the form

\[
c_b(\vec{q}) \cdot y_b(\vec{q}, t_0) = 0 ,
\]

then one has to demand that all \(C_{ab}\) and \(c_b\) coefficients be regular for \(\vec{q} \to 0\). This is the only technical assumption of the theorem \(^9\). If it is obeyed, then the two zero-momentum gauge modes can be promoted to physical, finite-momentum solutions, which are adiabatic by construction, and one of which turns out to have constant \(\zeta\) and \(\mathcal{R}\).

\(^9\)There is also an implicit assumption—used to write down the zero-momentum pure gauge solutions—that all background matter fields only depend on time, which is not obeyed in our case. However, this is easily fixed by performing the correct gauge transformation, which, following the notation of ref. \([8]\), in our case yields the pure gauge solution

\[
\Psi = H \varepsilon - \lambda , \quad \Phi = -\dot{\varepsilon} , \quad \pi = \lambda \bar{\varepsilon} , \quad \delta \rho = -\dot{\bar{\rho}} \varepsilon , \quad \delta p = -\dot{\bar{p}} \varepsilon .
\]
From our physical argument above, we see that the culprit for us is the anisotropic stress, which does not become negligible at long wavelengths. Indeed, comparing (4.117) with (4.115) we get schematically
\[ \delta \sigma \sim \frac{q}{q^2} \cdot \pi \]  
(4.121)
(we are neglecting factors of \( a(t) \), and corrections to (4.115) involving the metric, which do not change our conclusions.) Once plugged into eq. (4.116), this gives us an equation of motion that is not regular for \( q \to 0 \), thus violating Weinberg’s technical assumption. We get a similar singularity in the second equation of (4.116), since from \( T_{0i} \sim \pi^i \) we get a velocity potential
\[ \delta u \sim \frac{q}{q^2} \cdot \dot{\pi} \]  
(4.122)
Notice that we cannot reabsorb the annoying \( q^{-2} \) factors into a new \( \pi \) field, thus making \( \delta \sigma \) and \( \delta u \) regular for \( q \to 0 \), because one of the equations (4.118) is the equation of motion for \( \pi \) itself, which is local in real space, thus analytic in \( q \) in Fourier space. If we were to divide it by \( q^2 \), to write an evolution equation for the new \( \pi / q^2 \) field, we would introduce singularities there.

### 4.8 Reheating and Post-Inflationary Evolution

Since our scalar perturbations are not adiabatic, our predictions for post-inflationary correlation functions on large scales can in principle be affected by local physical processes happening at reheating. For this reason, we need to specify a model of reheating to end inflation.

Eventually we want our inflation to end and to be followed by a standard hot Big-Bang phase, that is, we want the universe to reheate and to become radiation dominated. In our case, this process can be thought of as a phase-transition from a solid state to a relativistic fluid state. The advantage of our language in dealing with such a transition is that it describes both solids and fluids in terms of the same long-distance degrees of freedom, our scalars \( \phi^I \). Only, the fluid action (2.6) enjoys (many) more symmetries. So, regardless of the microscopic dynamics that are actually responsible for the phase transition, at long distances and time scales reheating corresponds to some sort of symmetry enhancement of
our action.

In terms of our infrared degrees of freedom, what triggers reheating? In standard slow-roll inflation, it is the inflaton itself, when its time-dependent background field reaches a critical value. On the other hand, in the absence of perturbations, our $\phi^I$'s are exactly time-independent: $\langle \phi^I \rangle = x^I$. However the metric is not, and there are gauge-invariant combinations like our

$$X = g^{\mu\nu} \partial_\mu \phi^I \partial_\nu \phi^I,$$

or the energy density and pressure in eq. (4.14), that do depend on time. Usually we are used to solids turning into liquids—that is, melting—when the temperature exceeds a critical value. But we can also envisage a solid that ‘melts’ at zero temperature, when one of the physical quantities above goes past a critical value.

In our case, we need this zero-temperature melting to be associated with a substantial release of latent heat, so that the fluid we end up with is (very) hot. In our case, we need this zero-temperature melting to be associated with a substantial release of latent heat, so that the fluid we end up with is (very) hot. As far as we can tell, this does not violate any sacred principles of thermodynamics.

More explicitly, we will choose the quantity $B \equiv \det B^{IJ}$ (c.f. (2.5)) as our physical ‘clock’ — i.e. for large $B$ the action has the general structure (4.5), whereas for $B$ below a critical value $B_c$, the action has the more restricted form (2.6). There is a caveat: in the presence of fluctuations, the hypersurface defined by $B$ reaching its critical reheating value is different from that defined by other scalars (such as $[B^2]$, $[B^3]$...) reaching their critical reheating values. As a result, some of our predictions for cosmological correlation functions might be model-dependent. However since after reheating, in the hot fluid phase, the lowest-order action (2.6) depends unambiguously on $B$, it is thus natural, although not obviously necessary, to postulate that reheating is triggered by the value of $B$.

In addition, we assume that reheating is instantaneous, that is, that our solid/fluid phase transition happens in a time interval that is much shorter than the Hubble scale $H$, which is reasonable in principle, but not necessary. One can also consider much slower transitions,

---

10 Helium offers an example of such a phenomenon: at zero temperature one can turn liquid helium into a solid by raising the pressure beyond $\sim 25$ atm, and melt it back again by lowering the pressure below that value.

11 In the fluid phase, which is described by the action (2.6), the temperature is given by $T = -\frac{4F}{3\sqrt{\det F}}$. 

which in field space would correspond to replacing the sharp critical values we have been talking about for our observables, with much more continuous transition regions. All our physical considerations above apply unaltered.

A few comments on technical details are in order: Firstly, since it is the scalar quantity $B$ that plays the role of “clock” in our model, it is easiest to work in unitary gauge (UG), in which constant time slices correspond to surfaces of uniform $B$ (the properties of this gauge are worked out in Appendix $E$). Secondly, given the short duration of the phase transition, we can further postulate that the transition is smooth, in the sense that the dynamical metric perturbations and their time derivatives in UG are continuous across the transition.

To sum up, our model for reheating can be captured by the following statements:

- Inflation ends at $t_e$, with $a(t_e)^{-6} = B_e$, where $t$ denotes the time in UG.
- The matter content in the post inflationary era takes the form of a perfect fluid, which can be described by the effective action
  \begin{equation}
  S_{\text{fluid}} = \int d^4 x \sqrt{-g} \tilde{F}(B) \tag{4.124}
  \end{equation}
  where $B = \det g^{IJ}$ in UG. The change from one equation of state to the other is effectively instantaneous.
- Energy transfer from the solid phase to the fluid phase during this short reheating period is complete, and the normalization of $\tilde{F}$ is restricted in such a way that energy conservation is respected, i.e.
  \begin{equation}
  \rho_{\text{solid}} = -F(X,Y,Z) \bigg|_{t_e} = -\tilde{F} \left( a(t_e)^{-6} \right) = \rho_{\text{fluid}}. \tag{4.125}
  \end{equation}
  However, generally $p_{\text{solid}} \neq p_{\text{fluid}}$, since the equation of state has been changed. Consequently, even though the Hubble parameter $H$ remains continuous, $\dot{H}$ does not.
- Smoothness: the dynamical d.o.f.’s in unitary gauge—$A$ (or $\chi$), $C_i$, $D_{ij}$—as well as their first derivatives are continuous across the $t = t_e$ surface. The second derivatives will exhibit discontinuities since the equations of motions are altered due to the instantaneous change in the equation of state.

\footnote{See Appendix $E$ for the definition of these fields in terms of fluctuations of the metric.}
An immediate consequence following the smoothness requirement is that $\zeta$ transits continuously from solid phase to post-inflationary fluid phase, while $R$ does discontinuously. Indeed notice that in UG, $\zeta$ and $R$ are given by
\[
\zeta = \frac{A}{2}, \quad R = -\frac{H}{2H} \frac{\dot{A} - \dot{H} A / H}{1 - k^2 / 3a^2H} \quad (4.126)
\]
thus the discontinuity of $R$ stems from that of $\dot{H}$.

Given that the transition from the solid phase during inflation to the perfect fluid phase during the post-inflationary era occurs effectively instantaneously, we can compute various correlators (of $A$, $C_i$ and $D_{ij}$) at $t_e$ and use them as the initial conditions of the post-inflationary evolution. There are two subtleties:

1. How do we relate the correlators of quantities in UG to those in SFSG, which have been computed in Section 4.4 and Section 4.5.

2. Given that the super-horizon modes of $A$ and $D_{ij}$ are not adiabatic during inflation, they start the post-inflationary evolution with a non-vanishing first derivative in time. Although eventually these super-horizon modes become constant (time-independent) before reentering the horizon, a natural question to ask is how much the eventual constants could differ from the initial conditions these modes start with.

In order to address the first issue, let’s compute the scalar two- and three-point correlators: $\langle A_1(\tilde{t}_e) A_2(\tilde{t}_e) \rangle$ and $\langle A_1(\tilde{t}_e) A_2(\tilde{t}_e) A_3(\tilde{t}_e) \rangle$, where the “-” is to remind ourselves that we are using time in UG, and $A_i(\tilde{t})$ is shorthand for $A(\vec{k}_i, \tilde{t})$. Using the transformation rule from SFSG to UG, we have that
\[
\tilde{t} = t - \frac{1}{3H} \partial_i \pi^i(x) + O(\pi^2),
\]
and we can write
\[
A(x) = A^{(1)} + A^{(2)} + \ldots,
\]
where $A^{(1)}(x) = \frac{2}{3} \partial_i \pi^i(x) = 2\zeta$, $A^{(2)} \sim \partial \pi \partial \pi$, etc.

It follows immediately that, schematically,
\[
\langle A^2 \rangle = \langle A^{(1)} A^{(1)} \rangle + 2\langle A^{(1)} A^{(2)} \rangle + \ldots \sim 4\langle \zeta^2 \rangle + \langle (\partial \pi)^3 \rangle + \ldots \quad (4.127)
\]
The second term on the right can be neglected, since it is of higher order in the perturbative expansion. Likewise, we have

$$\langle A_1(\bar{t})A_2(\bar{t}) \rangle \simeq \langle A_1^{(1)}(\bar{t})A_2^{(1)}(\bar{t}) \rangle = 4\langle \zeta_1(\bar{t})\zeta_2(\bar{t}) \rangle \simeq 4\langle \zeta_1(t)\zeta_2(t) \rangle$$

(4.128)

where the last (approximate) equality is justified as long as the perturbative expansion (in fields) holds, since the difference between $t$ and $\bar{t}$ is of first order in the fields.

We can do the same for the 3-pt correlators:

$$\langle A_3 \rangle \simeq \langle A^{(1)}A^{(1)}A^{(1)} \rangle + 3\langle A^{(1)}A^{(1)}A^{(2)} \rangle \sim 8\langle \zeta^3 \rangle + \langle \zeta^2 \partial \pi \partial \pi \rangle.$$  

(4.129)

Notice that $\langle \zeta^2 \partial \pi \partial \pi \rangle \sim \langle \partial \pi \partial \pi \rangle^2 \sim O(\epsilon^{-2})$, while $\langle \zeta^3 \rangle \sim O(\epsilon^{-3})$, thus if we restrict ourselves to the leading order in slow-roll, this term can be safely neglected. It follows that

$$\langle A_1(\bar{t})A_2(\bar{t})A_3(\bar{t}) \rangle \simeq 8\langle \zeta_1(\bar{t})\zeta_2(\bar{t})\zeta_3(\bar{t}) \rangle$$

(4.130)

$$\simeq 8\langle \zeta_1(t)\zeta_2(t)\zeta_3(t) \rangle + \frac{8k_1}{3H}\langle \dot{\zeta}_1(t)\pi_{L,1}(t)\zeta_2(t)\zeta_3(t) \rangle + \text{perms.}$$

$$\simeq 8\langle \zeta_1(t)\zeta_2(t)\zeta_3(t) \rangle + \epsilon \cdot \frac{32c_{T}^2}{3}\langle \zeta_1(t)\zeta_1(t)\zeta_2(t)\zeta_3(t) \rangle + \text{perms.,}$$

where we have used that in the long wavelength limit $\dot{\zeta} \simeq \zeta' / a \simeq \frac{4}{3}c_T^2\epsilon H \zeta$. In the last line, the second term and its permutations are negligible at the leading order in slow roll, since $\epsilon \langle \zeta^4 \rangle \sim \epsilon \langle \zeta^2 \rangle^2 \sim O(\epsilon^{-1})$ while $\langle \zeta^3 \rangle \sim O(\epsilon^{-3})$.

So, as long as we focus only on the leading contribution in slow roll, the first issue mentioned above can be easily resolved: the 2-pt and 3-pt correlators of scalar perturbations in UG are related to those of (2 times) $\zeta$ in SFSG. Not surprisingly, similar relations for tensor perturbations hold if we apply the same logic.

As for the second issue. It can be shown in Chapter 2.4 (also in Ref. [10][12]) that during the post inflationary era, when the matter content of the universe is in the form of a perfect fluid, the scalar perturbation $A$ is adiabatic in the long wavelength limit, i.e. it is a constant at nonlinear level as long as it stays outside the horizon. However, unlike other inflationary models where there exists a conserved scalar mode in the long wavelength limit during inflation, the scalar perturbation $A$ in our model evolves slowly outside the horizon, in the sense that $A \simeq A^{(1)} = 2\zeta \propto (-\tau)^{\frac{4}{3}c_T^2 \epsilon}$. Therefore, after the rapid transition from solid phase to perfect fluid phase, rather than staying at its initial value, $A(t_e)$, it approaches its
eventual constant value. However the relative difference between the two is only of order $\epsilon$: the slow time-dependence of $A$ during inflation means that right after reheating the initial condition for the velocity is roughly $\dot{A}(t_e) \sim \epsilon \cdot HA(t_e)$. Since then, $\dot{A}$ decreases like $1/a^3$, thus making $A(t)$ approach its asymptotic value in a few Hubble times, during which $A(t)$ moves by $\sim \epsilon A(t_e)$. At the leading order in slow-roll, we can neglect this difference. Notice that this effect cannot change the tilts that we have computed: all modes of interest are outside the horizon during reheating and during the phase when $A$ relaxes to its asymptotic value. As a result, this small correction of order $\epsilon$ to the value of $A$ is the same for all modes, i.e., independent of $k$. By applying the same logic, we can reach the same conclusion for the transverse traceless tensor perturbation $D_{ij}$, which is not conserved in the long wavelength limit during inflation, but approaches an asymptotic value in the post-inflationary fluid phase in a similar manner as its scalar counterpart.

In conclusion, at the order we are working, we can take the correlation functions for $\zeta$ and $\gamma$ that we have computed during inflation in SFSG, evaluate them right before reheating, and obtain in this way good approximations to the corresponding correlation functions in UG in the post-inflationary phase. In particular, even though our scalar perturbations are not adiabatic during inflation, at reheating they get converted to adiabatic ones, with the same asymptotic constant value of $\zeta$ (up to $O(\epsilon)$ corrections) as they had at reheating.
Chapter 5

Conclusions

The main goal of this thesis is to demonstrate the power of the effective field theory approach beyond its traditional applications in particle physics. If we are only interested low energy phenomena, this approach allows us to be ignorant or agnostic about short distance physics; the only things we need to know are the effective low energy d.o.f.’s and symmetries, in addition to some observations from experiments. The former usually involves guesswork, while the latter fixes the undermined coefficients for us. We can apply this EFT approach to mundane systems like fluids as well as to a gravitational system; furthermore this idea guides us to construct inflation models of both theoretical and observational significance.

In chapter 2, we used this EFT approach to derive a low energy effective description of ordinary fluids. The IR degrees of freedom are identified as the volume elements’ positions—plus, if necessary, an extra scalar field standing for the $U(1)$ phase needed to model a conserved charge. Internal symmetries are imposed to ensure the homogeneity and isotropy of the ground state of fluids. The effective Lagrangian is organized as a derivative expansion, with the leading order corresponding to the non-dissipative sector of fluids, i.e. the perfect fluids. From the perfect fluid Lagrangian, one can recover the thermodynamical relations, with the understanding that these quantities like $T, \mu,...$ are intrinsic to each fluid volume element and are the consequence of some sort of ‘smearing’ process over the microscopic degrees of freedom of fluids. We also argued that an ordinary fluid free of vortex modes is classically equivalent to a zero-temperature superfluid.

To include the dissipative effects, we need to continue our construction of the fluid’s
EFT up to first order in derivatives. The beauty of this EFT approach lies in the fact that we can derive the linear dissipative effect without detailed knowledge of the dynamics of the underlying microscopic d.o.f. 's, which the IR modes can ‘lose’ energy to. More concretely, purely from symmetry arguments and the principles of EFT, we were able to derive that the coupling of hydrodynamical modes (i.e. the IR modes) to a generic thermalized sector that “lives in fluid” yields dissipation, with attenuation rates scaling as $k^2$. For fluids without conserved charges, the living-in-the-fluid requirement is strong enough to determine—via symmetry considerations—the precise structure of the interactions, thus allowing us to re-derive Kubo relations.

Having a Lagrangian description for perfect fluids enables us to compute straightforwardly the scattering amplitudes involving sound waves or vortex modes. From the 2 to 2 transverse scattering process, we found that the strong coupling scale associated with vortex modes vanishes. This is an inevitable consequence of the volume preserving diffeomorphism symmetry of the fluid system, Eqn. (2.4), which sets to zero the transverse speed of sound $c_T$. This puzzling feature seems to imply that our EFT description suffers from a strong coupling problem at arbitrarily low energy and is hence inconsistent at quantum level. On the other hand, we argued that this may just comes from a wrong identification of the vacuum state — the quantum fluctuations tend to restore the spontaneously broken symmetries. For the moment, we do not have a definite answer for the above puzzle and we will leave it to further work.

In the last section of Chapter 2, we focused on cosmological fluids. We have used the effective Lagrangian to extend the definition of curvature perturbation $\zeta$ to nonlinear orders and showed that it is conserved outside the horizon to all orders. Our way of constructing $\zeta$ is novel and has its own merits: 1) our proof did not rely on the assumption that the universe looks locally like an FRW patch (locally homogeneous and isotropic) on sufficiently large scales; and 2) we did not neglect the vector and tensor perturbations. Nevertheless, we showed that our definition of $\zeta$ agrees with that in Lyth et al. (2005) and Malik et al. (2004) up to a global reshuffling of the spatial coordinates.

In Chapter 3, we concentrated on gravity theories. From the EFT viewpoint, GR is unambiguously the low energy theory for a single massless spin-2 field. Any attempt to
modify GR in the infrared involves adding new degrees of freedom. The scalar tensor theory (with only one scalar field) is a natural starting point. In order to be considered as a qualified modified gravity model, first these theories must be consistent (for instance there should be no ghost excitations, and the propagating modes must travel at a subluminal speed). Meanwhile, they have to possess some screening mechanisms to shield the scalar force from detection, so that predictions will be in agreement with local gravity tests.

In particular, under very general conditions, we proved, two theorems for chameleon like theories, which limit their cosmological impact: i) the Compton wavelength of such a scalar can be at most Mpc at present cosmic density, which restricts its impact to non-linear scales; ii) the conformal factor relating Einstein- and Jordan-frame scale factors is essentially constant over the last Hubble time, which precludes the possibility of self-acceleration. These results imply that chameleon-like scalar fields have a negligible effect on the linear-scale growth history; theories that invoke a chameleon-like scalar to explain cosmic acceleration rely on a form of dark energy rather than a genuine modified gravity effect.

We also checked the stability of classical solutions of Galileon theory. Due to the complicated structure of the Galileon theory, it generally becomes extremely hard to prove the stability of a particular solution to the e.o.m. Instead, we introduced the concept of absolute stability, inspired by its cousin theory — the DGP model —, to circumvent this. By absolute stability of a theory we mean: if one can show that a field at a single point—like infinity for instance—in spacetime is stable, then stability of the field over the rest of spacetime is guaranteed for any positive energy source configuration. We found that the DGP parametrization ($c_4 = c_5 = 0$) is the only class of Galileon theories that is absolutely stable. Our analysis indicates that the DGP model is an exceptional choice among the large class of possible single field Galileon theories, and implies that if general solutions (non-spherically symmetric) exist, they may be unstable.

In chapter 4, we followed an EFT construction similar to treat of the fluid system, and reached an inflation model — solid inflation — with conceptually novel features that make it stand out as a radical alternative to the standard inflationary scenario. Our model differs drastically from more standard ones in its symmetry breaking pattern. In particular,
time-translations are not broken: there are physical “clocks”—i.e., time-dependent gauge-invariant observables,— but they inherit their time-evolution from the metric, not from the matter fields. As a result, the systematics of the EFT for the associated Goldstone excitations is completely different than the standard effective field theory of inflation. This has far reaching implications, some of which are directly observable. For instance, our model predicts (i) A nearly scale-invariant spectrum of adiabatic scalar perturbations, in agreement with observations. (ii) A nearly scale-invariant spectrum of tensor modes, with a slight blue tilt. (iii) The tensor-to-scalar ratio \( r \sim \epsilon c_L^5 \) ranges from somewhat smaller than in standard slow-roll inflation, for ultra-relativistic longitudinal phonons with \( c_L^2 \approx 1/3 \), to tiny, if they are non-relativistic. and (iv) A scalar three-point function with a novel shape—peaked in the squeezed limit, with non-trivial angular dependence on how the limit is approached—and a potentially very large amplitude, as big as \( f_{NL} \sim \frac{1}{\epsilon c_L^2} \). It would be interesting to run a dedicated numerical analysis of CMB data for our specific three-point function template, given its small overlaps with the more standard templates that have been considered so far.
Bibliography


Appendix A

Linear couplings of a $U(1)$ Goldstone

Consider a pair of complex scalar fields charged under a $U(1)$ global symmetry. For consistency of notation with our fluid case, let us denote them as $\phi$ and $\chi$. Then we have a Lagrangian of the sort

$$\mathcal{L} = \mathcal{L}_\phi[\phi] + \mathcal{L}_\chi[\chi, \phi], \quad (A.1)$$

invariant under the $U(1)$ transformation

$$\phi \to e^{i\alpha} \phi, \quad \chi \to e^{iq\alpha} \chi \quad (A.2)$$

(we are allowing for different charges for $\phi$ and $\chi$.) Notice that for the moment we are using a slightly different notation from the main text: we are including in $\mathcal{L}_\chi$ both the $\chi$-sector’s dynamics, and its interactions with the $\phi$ sector. For instance, $\mathcal{L}_\chi$ might contain interactions of the form

$$\mathcal{L}_\chi \supset \lambda(\phi^2 \chi^* \chi^2 + \text{h.c.}) \quad (A.3)$$

As in the case of spatial translations for our fluid, let us imagine now that this $U(1)$ symmetry is spontaneously broken by the vev of $\phi$: $\langle \phi \rangle = v$. Then, as it is standard, we can parameterize $\phi$ as

$$\phi = (v + \rho)e^{i\pi}, \quad (A.4)$$
APPENDIX A. LINEAR COUPLINGS OF A $U(1)$ GOLDSTONE

with $\pi$ being the Goldstone boson associated with the symmetry breaking, and $\rho$ the (generically) heavy radial excitation, which can be ignored at very low energies. For the sake of argument, let us thus set $\rho$ to zero from now on. Note that the symmetry is now realized non-linearly on $\pi$, i.e. $\pi \to \pi + \alpha$.

We can expand the action as

$$\mathcal{L} = \mathcal{L}_\phi[v e^{i\pi}] + \mathcal{L}_\chi[\chi, v e^{i\pi}]$$  \hspace{1cm} (A.5)

the same way we expanded $\phi^I = x^I + \pi^I$ for the fluid. However, in this parameterization of the fields it is not obvious that $\pi$ is derivatively coupled, as guaranteed from standard soft-pion theorems for the emission of a single soft $\pi$ quantum. For instance, from (A.3) we get a coupling

$$\mathcal{L}_\chi \supset \lambda v^2 (e^{i 2 q \pi} \chi^* e^{i 2 q \pi} + \text{h.c.}) ,$$  \hspace{1cm} (A.6)

which does not involve derivatives of $\pi$. This stems from a suboptimal choice of the field variables, and is easily remedied via a non-linear redefinition of the $\chi$ field, which as usual does not change the $S$-matrix:

$$\chi = \chi' e^{i q \pi} .$$  \hspace{1cm} (A.7)

Notice that the new $\chi'$ field is invariant under the $U(1)$ symmetry—the transformation of $\chi$ is now carried solely by the $e^{i q \pi}$ factor—and the action becomes:

$$\mathcal{L} = \mathcal{L}_\phi[v e^{i\pi}] + \mathcal{L}_\chi[\chi' e^{i q \pi}, v e^{i\pi}]$$  \hspace{1cm} (A.8)

Let’s focus on the $\mathcal{L}_\chi$ part. By the $U(1)$ symmetry—which now only acts on $\pi$—this action must be invariant under constant $\pi$ shifts, $\pi \to \pi + \alpha$. Then, interpreting the $\pi$ in $\mathcal{L}_\chi[\chi' e^{i q \pi}, v e^{i\pi}]$ as a weakly spacetime-dependent $U(1)$ transformation parameter, from Nöther’s theorem we get

$$\mathcal{L}_\chi[\chi' e^{i q \pi}, v e^{i\pi}] = \mathcal{L}_\chi[\chi', v] - \partial_\mu \pi J^\mu_\chi + \ldots ,$$  \hspace{1cm} (A.9)

where $J^\mu_\chi$ is the $\chi$-sector’s contribution to the $U(1)$ Nöther current, and we omitted terms with more $\pi$’s or more derivatives. In other words, at linear order in $\pi$ and at lowest order in the derivative expansion, the interaction between $\pi$ and the $\chi$ sector has to take the form

$$\mathcal{L}_{\text{int}} \simeq -\partial_\mu \pi J^\mu_\chi ,$$  \hspace{1cm} (A.10)
which is exactly the $U(1)$ analog of our eq. (2.62). Notice that, at this order, it does not matter whether we evaluate $J_\chi^\mu$ at $\chi$ or $\chi'$, since their difference is of first order in $\pi$. Notice also that we never really used that $\chi$ is scalar. Clearly our proof is completely general and holds for any set of charged fields $\chi$, of any spin.

As an alternative, quicker derivation of the same result, we can go back to eq. (A.5) and use the following trick (a version of Stückelberg’s trick): We promote the global symmetry to a gauge symmetry by introducing an auxiliary gauge field $A_\mu$ which transforms as $A_\mu \rightarrow A_\mu - \partial_\mu \alpha(x)$, and replace standard derivatives by covariant ones. Expanding the action to linear order in $A_\mu$ we have

$$L[\phi, \chi, A_\mu] = L[\phi, \chi] + J_\mu A^\mu + \mathcal{O}(A^2),$$

where $J_\mu[\phi, \chi]$ is the conserved current (in the absence of $A^\mu$) for the $U(1)$ global charge.\footnote{Note we included all the gauge-field dependence explicitly (in the $\mathcal{O}(A^2)$ terms), so that there is no $A^\mu$ in $J^\mu$, which would be necessary to make it a gauge invariant expression.} Since we promoted this symmetry to a gauge transformation we are guaranteed that $\pi$ disappears from the action, because it can be absorbed into $A_\mu$ by choosing $\alpha = -\pi$. This means that, to linear order in $\pi$, we wind up with the coupling\footnote{Notice that $J_\mu$ in Eq. (A.11) does depend on $\pi$, e.g. $J_\mu \supset f \partial_\mu \pi$, for some 'decay constant’ $f$. However this dependence, and the one stemming from $L[\phi, \chi]$, is ultimately absorbed into $A^\mu$ including the $\mathcal{O}(A^2)$ piece. That is why only $J_\chi^\mu$ enters in the coupling to linear order.}

$$J_\chi^\mu \partial_\mu \pi,$$

where $J_\chi^\mu$ is the $\chi$-dependent component of the full current at zeroth order in $\pi$.\footnote{The alert reader may have already recognized this is the way longitudinal gauge bosons couple to matter.} Hence we conclude that, at leading order in the perturbations, the Goldstone boson couples to the $\pi$-independent part of the current associated with the broken symmetry.

The introduction of $A_\mu$ is equivalent to working in the so called unitary gauge, where the Goldstones are set to zero and their interactions are encoded in the gauge field. The previous analysis suggests that one could perform similar manipulations in the case of our fluid, where the Goldstone fields $\pi^I$ are associated with the breaking of spatial translations.
Now, to go to the unitary gauge we must introduce the gauge field associated with spatial translations, namely the metric perturbation $h_{\mu I}$, and the broken generators are the $T^{\mu I}$ components of the stress energy tensor. Hence, the coupling must read: $T^{\mu I} \chi h_{\mu I}$. To introduce the pions we do as before, which in our case entails schematically

$$h_{\mu I} = \partial_{(\mu \alpha_I)} \to \partial_{(\mu \pi_I)}. \quad (A.13)$$

This viewpoint might prove useful in extending our results to non-linear order.
Appendix B

Phase Space for Phonons

We are mostly interested in a two-particle final state, possibly with two independent propagation speeds. The infinitesimal phase space is

$$d\Pi_2 = (2\pi)^4 \delta^3(\vec{P} - \vec{q}_1 - \vec{q}_2) \delta(E - E_1 - E_2) \cdot \frac{d^3q_1}{(2\pi)^3} \frac{d^3q_2}{(2\pi)^3} \frac{1}{2E_1} \frac{1}{2E_2}, \quad (B.1)$$

where $E$ and $\vec{P}$ are the total energy and momentum. The integral in $\vec{q}_2$ eliminates the momentum-conservation delta-function. Then we are left with

$$d\Pi_2 = \frac{d\Omega}{(2\pi)^2} q_1^2 q_2 \frac{1}{2E_1} \frac{1}{2E_2} \delta(E - E_1 - E_2), \quad (B.2)$$

with the understanding that $E_2$ be evaluated at $\vec{q}_2 = \vec{P} - \vec{q}_1$. We have

$$\delta(E - E_1 - E_2) = \frac{\delta(q_1 - \vec{q}_1)}{\partial E_1 \partial q_1 + \partial E_2 \partial q_2} \quad (B.3)$$

and

$$\frac{\partial q_2}{\partial q_1} \equiv \frac{\partial |\vec{P} - \vec{q}_1|}{\partial q_1} = \frac{q_1 - P \cos \theta}{q_2}, \quad (B.4)$$

where $\theta$ is the angle between $\vec{q}_1$ and $\vec{P}$. On the other hand, the derivatives of the energies w.r.t. the corresponding momenta are the particles’ group velocities. Integrating over $q_1$ we thus get

$$d\Pi_2 = \frac{d\Omega}{16\pi^2} \frac{q_1^2 q_2}{E_1 E_2} \frac{1}{|c_1 q_2 + c_2 q_1 - c_2 P \cos \theta|} \quad (B.5)$$

For a linear dispersion law like in our case, $E_a = c_a q_a$, we finally have

$$d\Pi_2 = \frac{d\Omega}{16\pi^2} \frac{1}{c_1 c_2} \frac{q_1}{|c_1 q_2 + c_2 q_1 - c_2 P \cos \theta|}. \quad (B.6)$$

In special circumstances there are further simplifications:
APPENDIX B. PHASE SPACE FOR PHONONS

i) For scattering processes at zero total momentum, we can set \( P = 0 \) and \( q_1 = q_2 \). We get

\[
d\Pi_2 = \frac{d\Omega}{16\pi^2} \cdot \frac{1}{c_1 c_2 (c_1 + c_2)} \quad (\vec{P} = 0).
\] (B.7)

ii) For decay processes at finite total \( \vec{P} \), but when one of the final particles is much slower than the other, barring an hierarchy between \( q_1 \) and \( q_2 \) we have

\[
d\Pi_2 \simeq \frac{d\Omega}{16\pi^2} \frac{1}{c_1^2 c_2} \frac{q_1}{q_2} \quad (c_2 \ll c_1).
\] (B.8)

Of course the ratio \( q_1/q_2 \) depends non-trivially on the angle \( \theta \) we are supposed to integrate over—which we can take to be the angle between \( \vec{q}_1 \) and \( \vec{P} \). We have:

\[
\frac{q_1}{q_2} \simeq \frac{1}{2 \sin \theta/2} \quad (c_2 \ll c_1, \vec{P} \neq 0).
\] (B.9)

Overall we thus get

\[
d\Pi_2 \simeq \frac{d\Omega}{32\pi^2} \frac{1}{c_1^2 c_2} \frac{1}{\sin \theta/2} \quad (c_2 \ll c_1, \vec{P} \neq 0).
\] (B.10)

Notice that this is regular at \( \theta = 0 \), thus making our ‘barring an hierarchy . . . ’ approximation under control. That is, eq. (B.10) is the correct phase-space element at lowest order in \( c_2/c_1 \).
Appendix C

Expressions for $Z^\mu_\nu$ in the Galileon Theory

Following equation (3.66), it is straightforward to calculate the various $Z^\mu_\nu$’s,

$$Z^\mu_\nu = \sum_{i=2}^{5} c_i Z^{(i)} \mu_\nu$$

with the $Z^{(i)}$’s given in their matrix form by

$$Z^{(2)} = -\delta$$
$$Z^{(3)} = -2\text{Tr}[\Pi] \delta + 2\Pi$$
$$Z^{(4)} = -3(\text{Tr}[\Pi])^2 \delta + 3\text{Tr}[\Pi \cdot \Pi] \delta + 6\text{Tr}[\Pi][\Pi] \Pi - 6\Pi \cdot \Pi$$
$$Z^{(5)} = -4(\text{Tr}[\Pi])^3 \delta + 12\text{Tr}[\Pi][\Pi][\Pi] \delta + 12(\text{Tr}[\Pi])^2 \Pi$$
$$-8\text{Tr}[\Pi \cdot \Pi \cdot \Pi] \delta - 24\text{Tr}[\Pi][\Pi \cdot \Pi] \delta - 12\text{Tr}[\Pi \cdot \Pi][\Pi] \Pi + 24\Pi \cdot \Pi \cdot \Pi$$

(C.1)

where $\delta$ denotes the $4 \times 4$ identity matrix, $\Pi$ the matrix $\Pi^\mu_\nu \equiv \partial^\mu \partial_\nu \pi$, and $\text{Tr}$ and “·” stand for the trace and matrix multiplication respectively.
Appendix D

Further Details of the Galileon Stability Analysis

We will simplify equation (3.100) using some algebraic tricks. In analyzing the conditions for the $Z_1 = 0$ region we begin with the second condition in (3.100)

$$\Gamma = 9\alpha_1^2 + 6 - 18\alpha_1\alpha_3 - 3\alpha_3^2 + 8\alpha_1\alpha_3^3 \geq 0 \quad (D.1)$$

setting $\Gamma = 0$ and solving for $\alpha_1$

$$\alpha_1 = \alpha_3 - \frac{4}{9}\alpha_3^3 - \frac{\sqrt{2}}{9}(2\alpha_3^2 - 3)^{3/2} \equiv \alpha_- \quad (D.2)$$

or

$$\alpha_1 = \alpha_3 - \frac{4}{9}\alpha_3^3 + \frac{\sqrt{2}}{9}(2\alpha_3^2 - 3)^{3/2} \equiv \alpha_+ \quad (D.3)$$

Note, that since $2\alpha_3^2 - 3 > 0$ both of these solutions are real. Thus, we can rewrite the second condition as

$$\Gamma \geq 0 \iff \alpha_1 \leq \alpha_- \text{ or } \alpha_1 \geq \alpha_+ \quad (D.4)$$

Now, let’s consider the third constraint in (3.100)

$$((9\alpha_1 - 9\alpha_3 + 4\alpha_3^3)k_1\sqrt{c_4} + (-3 + 3\alpha_1\alpha_3 + \alpha_3^2)) \geq \sqrt{\Gamma}|(3k_1\sqrt{c_4} + \alpha_3)| \quad (D.5)$$

$$\iff ((9\alpha_1 - 9\alpha_3 + 4\alpha_3^3)k_1\sqrt{c_4} + (-3 + 3\alpha_1\alpha_3 + \alpha_3^2)) \geq 0 \text{ and } \quad (D.6)$$

$$((9\alpha_1 - 9\alpha_3 + 4\alpha_3^3)k_1\sqrt{c_4} + (-3 + 3\alpha_1\alpha_3 + \alpha_3^2))^2 \geq \Gamma(3k_1\sqrt{c_4} + \alpha_3)^2 \quad (D.7)$$

Notice that when $\alpha_1 \leq \alpha_-$, the prefactor to $k_1$, $(9\alpha_1 - 9\alpha_3 + 4\alpha_3^3) \leq -\sqrt{2}(2\alpha_3^2 - 3)^{3/2} < 0$. Similarly, when $\alpha_1 \geq \alpha_+$ the prefactor $\geq +\sqrt{2}(2\alpha_3^2 - 3)^{3/2} > 0$. Thus, we can combine the
inequalities (D.6) and (D.4) as
\[ \alpha_1 \leq \alpha_- \quad \text{and} \quad k_1 \sqrt{c_4} \leq K(\alpha_1, \alpha_3) \quad \text{or} \quad \alpha_1 \geq \alpha_+ \quad \text{and} \quad k_1 \sqrt{c_4} \geq K(\alpha_1, \alpha_3) \quad (D.8) \]
where
\[ K(\alpha_1, \alpha_3) \equiv (-3 + 3\alpha_1 \alpha_3 + \alpha_3^2) \left/ (9\alpha_1 - 9\alpha_3 + 4\alpha_3^2) \right. = -\frac{\alpha_3}{3} + \frac{(2\alpha_3^2 - 3)^2}{3(9\alpha_1 - 9\alpha_3 + 4\alpha_3^2)} \quad (D.9) \]

Now, consider (D.7) which can be rewritten as
\[ (3 - 2\alpha_3^2)^2 \left\{ 1 - 2\alpha_1 \alpha_3 + k_1 \sqrt{c_4}(-6\alpha_1 + 2\alpha_3) + (k_1 \sqrt{c_4})^2(4\alpha_3^2 - 6) \right\} \geq 0 \quad \Leftrightarrow \quad (D.10) \]
\[ 1 - 2\alpha_1 \alpha_3 + k_1 \sqrt{c_4}(-6\alpha_1 + 2\alpha_3) + (k_1 \sqrt{c_4})^2(4\alpha_3^2 - 6) \geq 0 \quad \Leftrightarrow \quad (D.11) \]
\[ k_1 \leq \frac{1}{\sqrt{c_4}} \frac{3\alpha_1 - \alpha_3 - \sqrt{T}}{4\alpha_3^2 - 6} = k_2(\alpha_1, \alpha_3) \quad \text{or} \quad k_1 \geq \frac{1}{\sqrt{c_4}} \frac{3\alpha_1 - \alpha_3 + \sqrt{T}}{4\alpha_3^2 - 6} = k_3(\alpha_1, \alpha_3) \quad (D.12) \]

In the last line, (3.94) was used. Let’s compare (D.12) with our previous results (D.8). There are two regions we need to concern ourselves with: \( \alpha_1 \leq \alpha_- \) and \( \alpha_1 \geq \alpha_+ \). Amazingly, we find the combined result is simply
\[ \alpha_1 \leq \alpha_- \quad \text{and} \quad k_1 \sqrt{c_4} \leq \sqrt{c_4} k_2(\alpha_1, \alpha_3) \quad \text{or} \quad \alpha_1 \geq \alpha_+ \quad \text{and} \quad k_1 \sqrt{c_4} \geq \sqrt{c_4} k_3(\alpha_1, \alpha_3) \quad (D.13) \]
due to the observation that
\[ \sqrt{c_4} k_2(\alpha_1, \alpha_3) \leq \sqrt{c_4} k_2(\alpha_-, \alpha_3) = K(\alpha_-, \alpha_3) \leq K(\alpha_1, \alpha_3) \quad \text{for} \ \alpha_1 \leq \alpha_- \quad (D.14) \]
\[ \sqrt{c_4} k_3(\alpha_1, \alpha_3) \geq \sqrt{c_4} k_3(\alpha_+, \alpha_3) = K(\alpha_+, \alpha_3) \geq K(\alpha_1, \alpha_3) \quad \text{for} \ \alpha_1 \geq \alpha_+ \quad (D.15) \]

In order to have guaranteed stability we need the conditions contained in (D.13) and the first condition of (3.100), the \( \alpha_5 \) dependent one, to yield the null set. Equivalently, explicitly separating the two regions, we may write
- If \( \alpha_1 \leq \alpha_- \), for any \( k_1 \sqrt{c_4} \leq \sqrt{c_4} k_2(\alpha_1, \alpha_3) \), \( k_1 \sqrt{c_4} f(\alpha_1, \alpha_3, \alpha_5) < 0 \)
- If \( \alpha_1 \geq \alpha_+ \), for any \( k_1 \sqrt{c_4} \geq \sqrt{c_4} k_3(\alpha_1, \alpha_3) \), \( k_1 \sqrt{c_4} f(\alpha_1, \alpha_3, \alpha_5) < 0 \)

Where we have defined \( f(\alpha_1, \alpha_3, \alpha_5) \equiv (9 - 12\alpha_3 \alpha_5)\alpha_1 - 6\alpha_3 + 2\alpha_3^3 + 6\alpha_5 \). Let’s investigate when these conditions are satisfied – when we are guaranteed classical stability.

**Instability for** \( \alpha_1 \leq \alpha_- \)
APPENDIX D. FURTHER DETAILS OF THE GALILEON STABILITY ANALYSIS

As we are interested in the region where $\alpha_3 > \sqrt{3/2}$ and $\alpha_5 < 0$, we see that $(9 - 12\alpha_3\alpha_5) > 0$ and thus

$$f(\alpha_1, \alpha_3, \alpha_5) \leq f(\alpha_-, \alpha_3, \alpha_5)$$

$$= (2\alpha_3^2 - 3) \left\{-\alpha_3 - \sqrt{4\alpha_3^2 - 6 + \frac{\alpha_5}{3}(4\alpha_3\sqrt{4\alpha_3^2 - 6} + 8\alpha_3^2 - 6)} \right\}$$

$$< 0 \quad \text{for any } \alpha_5 < 0 \text{ and } \alpha_3 > \sqrt{3/2}$$

So, for $\alpha_1 \in (-\infty, \alpha_-]$, $f(\alpha_1, \alpha_3, \alpha_5) < 0$ and $k_2(\alpha_1, \alpha_3) < 0$, from which we can see that there always exists some $k_1\sqrt{c_4}$ such that $k_1\sqrt{c_4}f(\alpha_1, \alpha_3, \alpha_5)$ is positive. Thus, for $\alpha_1 \in (-\infty, \alpha_-]$, the condition $(\ref{condition})$ is not a null set and the corresponding parameter choice is not a stable one.

**Instability for $\alpha_1 \geq \alpha_+$**

Note that for this region $k_3(\alpha_1, \alpha_3)$ is a monotonically increasing function of $\alpha_1$

$$k_3(\alpha_1, \alpha_3) \geq k_3(\alpha_+, \alpha_3) = \frac{1}{3} \left(-\alpha_3 + \sqrt{\alpha_3^2 - \frac{3}{2}}\right)$$

and that $k_3(\alpha_+, \alpha_3) < 0$. There exists an $\alpha_*$ such that $k_3(\alpha_*, \alpha_3) = 0$. In fact, $\alpha_* = 1/2\alpha_3$. 

If $\alpha_+ \leq \alpha_1 \leq \alpha_*$, then $k_3(\alpha_+, \alpha_3) \leq k_3(\alpha_1, \alpha_3) \leq k_3(\alpha_*, \alpha_3) = 0$ and therefore the $k_1\sqrt{c_4}$'s, satisfying $k_1\sqrt{c_4} \geq \sqrt{c_4}k_3(\alpha_1, \alpha_3)$, could be either positive or negative and so be $k_1\sqrt{c_4}f(\alpha_1, \alpha_3, \alpha_5)$, regardless of the sign of $f(\alpha_1, \alpha_3, \alpha_5)$. Once again, the condition $(\ref{condition})$ is not a null set and the region $\alpha_1 \in [\alpha_+, \alpha_*]$ is not stable.

If $\alpha_1 \geq \alpha_*$, then $k_3(\alpha_1, \alpha_3) \geq k_3(\alpha_*, \alpha_3) = 0$. Since

$$f(\alpha_1, \alpha_3, \alpha_5) \geq f(\alpha_*, \alpha_3, \alpha_5) = \frac{(2\alpha_3^2 - 3)^2}{2\alpha_3} > 0$$

so it follows that for any $k_1\sqrt{c_4} (\geq \sqrt{c_4}k_3(\alpha_1, \alpha_3)), k_1\sqrt{c_4}f(\alpha_1, \alpha_3, \alpha_5)$ is positive. By the same argument above, we see that $\alpha_1 \geq \alpha_*$ is not a stable region.

In summary, after a long and torturous process we see that the only stable region (i.e. the choice of parameter making condition $(3.100)$ a null set) is determined by $\Gamma < 0$, or in other words

$$\alpha_- = \alpha_3 - \frac{4}{9}\alpha_3^3 - \frac{\sqrt{2}}{9}(2\alpha_3^2 - 3)^{3/2} < \alpha_1 < \alpha_3 - \frac{4}{9}\alpha_3^3 + \frac{\sqrt{2}}{9}(2\alpha_3^2 - 3)^{3/2} = \alpha_+$$
In performing calculations throughout this paper two gauge choices are particularly useful:

- Spatially Flat Slicing Gauge (SFSG) is defined by setting to zero the scalar and vector perturbations in $g_{ij}$, i.e. by imposing

$$g_{ij} = a(t)^2 \exp \gamma_{ij}, \quad (E.1)$$

where $\gamma_{ij}$ denotes the transverse traceless tensor mode, satisfying

$$\gamma_{ii} = 0, \quad \partial_i \gamma_{ij} = 0. \quad (E.2)$$

Then the fluctuations in our $\phi^I$ scalars are unconstrained:

$$\phi^I = x^I + \pi^I. \quad (E.3)$$

The three $\pi^I(x)$ fields can be split into a transverse vector and longitudinal scalar as in (??), according to their transformation properties under the residual rotation group. This gauge choice is of particular convenience for computations of the two- and three-point functions because in the demixing (with gravity) limit the $\pi$ Lagrangian will contain all the scalar (or longitudinal) and vector (or transverse) degrees of freedom.
- Unitary Gauge (UG) is defined by setting to zero the fluctuations in the $\phi^I$ fields and in the “clock” field:

$$\phi^I = x^I, \quad \det(B^{IJ}) = a(t)^{-6}. \quad (E.4)$$

Then the spatial metric is unconstrained. And can be parameterized in general as

$$g_{ij} = a(t)^2 \exp(A \delta_{ij} + \partial_i \partial_j \chi + \partial_i C_j + \partial_j C_i + D_{ij}), \quad (E.5)$$

where $\partial_i C_i = 0$ and $D_{ij}$ is transverse-traceless.

From the above form of the metric it seems that in UG there are too many degrees of freedom; there is an extra scalar in addition to the scalar, transverse vector, and transverse traceless tensor that we expect. However, when the metric is expressed in terms of the ADM parameters defined by (2.100) (2.101), the second condition in (E.4) can be rewritten as

$$3A + \nabla^2 \chi = \log(1 - N_i N^i/N^2). \quad (E.6)$$

As $N^i$ and $N$ can be expressed in terms of $A$, $\chi$, $C_i$, and $D_{ij}$ by solving the constraint equations given by (4.42), (4.43) we can see that (E.6) implies that the two scalar functions $A(x)$ and $\chi(x)$ are not independent in UG. Hence the dynamical d.o.f. in question can be chosen to be $A(x)$ (the scalar mode), $C_i(x)$ (the transverse vector mode), and $D_{ij}$ (the transverse traceless tensor mode). The number of which matches our physical intuition and properly agrees with SFSG. UG is particularly useful in following our degrees of freedom through the reheating surface.

As we find it convenient to calculate correlation functions in SFSG, and yet utilize UG to most easily describe the surface of sudden reheating, we need to develop the transformation rules to go from one gauge to the other. Let’s denote by $\{x^\mu\}$ the coordinate system in SFSG and by $\{\bar{x}^\mu\}$ that in UG. A gauge transformation relating SFSG to UG is given by $\bar{x}^\mu = x^\mu + \xi^\mu(x)$, where

$$\xi^0(x) = -\frac{1}{3H} \partial_i \pi^i(x) + O(\pi^2), \quad \xi^I(x) = \pi^I(x) + O(\pi^2) \quad (E.7)$$

and the scalar perturbations are related by

$$A = \frac{2}{3} \partial_i \pi^i + O(\pi^2). \quad (E.8)$$
Appendix F

Full trilinear action in SFSG for Solid Inflation

Expanding the Lagrangian to third order in fluctuations about the FRW background in SFSG we have after a straightforward but lengthy computation\footnote{Since we are after the three-point function for scalar perturbations, we ignore the interaction between the tensor mode $\gamma$ and the $\pi$ fields.}

$$L^{(3)} = a(t)^3 \left\{ 3M_{Pl}^2 H^2 \delta N^3 + 2M_{Pl}^2 H \partial_i N^i \delta N^2 + 2M_{Pl}^2 \dot{H} a^2 N^j \partial_j \pi^i (\dot{\pi}^i - N^i) ight. $$ 
$$\left. - M_{Pl}^2 \delta N \left( \frac{1}{4} \partial_i N^j \partial_j N^i + \frac{1}{2} \partial_i N^i \partial_j N^j - \frac{1}{2} (\partial_i N^i)^2 \right) \right. $$
$$\left. - \delta N M_{Pl}^2 \dot{H} \left( - a^2 (\dot{\pi}^i - N^i)^2 - c_T^2 [\Pi^T \Pi] + (1 - c_T^2) [\Pi^2] - (1 + c_L^2 - 2c_T^2) [\Pi^3] \right) \right. $$
$$+ [\Pi^3] \left( \frac{4}{3} F_{XX} a^{-6} - \frac{3}{2} F_{XZ} + F_{XY} \right) a^{-2} + \frac{64}{27} F_Z + \frac{16}{81} F_Y $$
$$+ [\Pi][\Pi^2] \left( \frac{1}{2} (F_{XZ} + F_{XY}) a^{-2} - \frac{4}{3} F_Z - \frac{8}{27} F_Y \right) $$
$$+ [\Pi][\Pi^T \Pi] \left( 2F_{XX} a^{-4} + \frac{4}{3} (F_{XZ} + F_{XY}) a^{-2} - \frac{16}{27} F_Z - \frac{4}{3} F_Y \right) $$
$$+ \frac{2}{27} F_Z [\Pi^2] + \left( \frac{2}{3} F_Z + \frac{4}{3} F_Y \right) [\Pi^T \Pi \Pi] - \frac{4}{3} (F_Y + F_Z) a^{-2} (\dot{\pi}^i - N^i) \partial_i \pi^j (\dot{\pi}^j - N^j) $$
$$+ a^2 [\Pi] (\dot{\pi}^i - N^i)^2 \left( - 2F_{XX} a^{-4} + \frac{4}{27} (F_Y + F_Z) \right) \right\} , \tag{F.1}$$

where $\Pi$ denotes the $3 \times 3$ matrix $\partial_i \pi^j$ and $[\cdots]$ indicates the trace; for instance, $[\Pi^T \Pi \Pi] \equiv \partial_j \pi^i \partial_j \pi^k \partial_k \pi^i$. 
APPENDIX F. FULL TRILINEAR ACTION IN SFSG FOR SOLID INFLATION

Now, as discussed in Section 4.3 one only needs to solve the constraint equations
\[ \delta S/\delta N = 0 \quad \text{and} \quad \delta S/\delta N^i = 0 \]
to linear order in perturbations. We therefore don’t need to worry about terms in \( N, N^i \) that are quadratic in the fluctuations contributing to the cubic Lagrangian. The solutions to these equations are given by (4.49), (4.50), and (4.51).

In particular, we are interested in two separate limits for computing the three-point function. The first is the de-mixing regime where \( k \gg aH\epsilon^{1/2} \). In this limit, to lowest order in slow roll, we can effectively set \( \delta N \) and \( N_L \) to zero. Furthermore, note that all terms that are not of the form \( \Pi^3 \) (like the final \( [\Pi]\pi^2 \) term) are explicitly suppressed by slow roll parameters. We are left with
\[
L_3 = a(t)^3 M_{Pl}^2 H^2 F_Y \left\{ \frac{7}{8\pi} (\partial \pi)^3 - \frac{1}{9} \partial \pi \partial_j \pi^k \partial_k \pi^j - \frac{4}{9} \partial \pi \partial_j \pi^k \partial_j \pi^k + \frac{2}{3} \partial_j \pi^i \partial_j \pi^k \partial_k \pi^i \right\},
\]
where we neglected boundary terms. We have freely used the Friedmann equations (4.18), various definitions of slow roll parameters, and the the total derivative
\[
\det(\partial \pi) = \frac{1}{6} \epsilon_{ijk} \epsilon_{lmn} \partial^i \partial^j \pi^j \partial^m \pi^m \pi^n = \frac{1}{6} \left( [\partial \pi]^3 - 3[\partial \pi][\partial \pi^2] + 2[\partial \pi^3] \right)
\]
(which is a total derivative because of the \( \epsilon \)-tensor structure).

The second limit is in the strong mixing (with gravity) regime. This occurs when \( k \ll aH\epsilon^{1/2} \). In this limit we can write to lowest order in slow roll
\[
\delta N = \frac{k}{a} \frac{d}{d\tau} \left( \frac{\pi_L}{H} \right) \simeq k\epsilon \pi_L,
\]
\[
N_L = \dot{\pi}_L, \quad (F.5)
\]
\[
N^i_T = \dot{\pi}^i_T, \quad (F.6)
\]
where we have estimated the time dependence of \( \pi \) via the explicit classical solution to the first order equation of motion given by (4.99a). When we insert the above expressions into the full cubic Lagrangian we see that all the terms involving these auxiliary fields are going to vanish, as the recurring combination \( (\dot{\pi}^i - N^i) \) vanishes to lowest order in slow roll and \( \delta N \) is explicitly of order \( \epsilon(\partial \pi) \). So, surprisingly, we see that to lowest order in slow roll we recover the same expression \( (F.2) \) for the cubic Lagrangian in the strong mixing limit. This is a convenient fact, as it allows us to effectively use the same expression for the cubic interactions during the whole inflationary phase in our calculation of the three-point function in Section 4.5.
Appendix G

Time Dependence of Background Quantities in Solid Inflation

In order to solve the classical equations of motion for scalar and tensor perturbations, we need know the explicit time dependence of quantities such as $a(\tau)$, $H(\tau)$, $\epsilon(\tau)$, \ldots; the goal of this section is obtaining this time dependence. For the computations we are interested in, it suffices to derive these temporal functions up to the first order in slow roll. To make the notation lighter, we will mostly drop the $\tau$ argument: $a(\tau) \to a$, etc. Primes will denote derivatives with respect to $\tau$.

Recall the definition of the first slow roll parameter $\epsilon$, (4.17), and rewrite it as

$$
\epsilon = -\frac{H'}{aH^2} = \frac{d}{d\tau} \left( \frac{1}{aH} \right) + 1.
$$

Integrating the above equation once and choosing some suitable additive constant $\Box$ one has

$$
\frac{1}{aH} = - (1 - \epsilon_c) \tau + O(\epsilon^2)
$$

where the subscript “c” denotes evaluation at some reference conformal time $\tau_c$, which we find most convenient to choose to be the (conformal) time when the longest mode of observational relevance today exists the horizon, i.e. $|c_{L,c}h_{\min}\tau_c| \simeq |c_{L,c}\tau_cH_{\text{today}}| = 1$.

\[1\] The integration constant is chosen by demanding $a(\tau) \gg a(\tau_c)$, for $\tau/\tau_c \to 0$. 

APPENDIX G. TIME DEPENDENCE OF BACKGROUND QUANTITIES IN SOLID INFLATION

The reason $\epsilon(\tau)$ is being treated as a constant in integration is that the higher order terms in the Taylor series of $\epsilon(\tau) = \epsilon(\tau_c) + \epsilon'(\tau_c)(\tau - \tau_c) + \ldots$ are suppressed by powers of slow roll, for instance $\epsilon'(\tau_c)\tau_c \sim O(\epsilon^2)$. Of course this also depends on the choice of reference time; we don’t want the perturbative expansion in slow roll to be contaminated by large values of $\tau/\tau_c - 1$. Since $|\tau(t)|$ is a decreasing function during inflation, and $+\infty > -\tau > 0$, we ought to choose early times (like $\tau_c$) as the reference.

Using the definition of the Hubble parameter, we can extract the time dependence of the scale factor $a(\tau)$ from the above equation (G.2):

$$aH = \frac{a'}{a} = -\frac{1 + \epsilon_c}{\tau} \implies a(\tau) = a_c \left(\frac{\tau}{\tau_c}\right)^{-1-\epsilon_c} + O(\epsilon^2) \quad (G.3)$$

Furthermore, we obtain

$$H(\tau) = \frac{a'}{a^2} = -\frac{1 + \epsilon_c}{a_c \tau_c} \left(\frac{\tau}{\tau_c}\right)^{\epsilon_c} + O(\epsilon^2). \quad (G.4)$$

Finally, the time dependence of $\epsilon$, $c_L$ and $c_T$ can be revealed by invoking the definitions of other slow roll parameters. For instance,

$$\frac{\epsilon'}{\epsilon} = aH\eta = -\frac{\eta}{\tau} + O(\epsilon^2) \implies \epsilon(\tau) = \epsilon_c \left(\frac{\tau}{\tau_c}\right)^{-\eta_c} + O(\epsilon^3). \quad (G.5)$$

Similarly we obtain

$$c_L(\tau) = c_{L,c} \left(\frac{\tau}{\tau_c}\right)^{-s_c} + O(\epsilon^2), \quad c_L(\tau) = c_{T,c} \left(\frac{\tau}{\tau_c}\right)^{-u_c} + O(\epsilon^2). \quad (G.6)$$

Notice that, because of the all-order relation between $c_T^2$ and $c_L^2$ of footnote 4, $c_{T,c}$ and $u_c$ are not independent parameters—they can be expressed in terms of $c_{L,c}$ and of the slow-roll parameters. The equations (G.3)–(G.6) are frequently used in solving the classical equations of motion.