GGI lectures

Cosmological tests of fundamental Physics: Harvesting the non-Gaussian Universe

P. Daniel Meerburg, Will Coulton

March 18-22, 2024

Contents

Lecture 1: Introduction

Before we start, let us pause and consider the topic of this lecture series, 'Cosmological tests of fundamental physics'. The organisers have asked to talk about this topic and I have decided to stick with it :-). However, one thing I think we should discuss is the 'Fundamental Physics' in the description of this lecture series. Admittedly, the notion that I (or colleagues in our field) aim to either develop a fundamental theory or test such a theory is pretty common. Seminars, colloquiums and research proposals often include the word 'fundamental'. The question is, what do we imply when we say 'fundamental physics'? Is not all physics fundamental? And who decides whether it is in the first place? If I were Isaac in 1687, I could have argued that I was working on fundamental physics. Similar if I were James Clarke in 1861. So at least is seems that whether physics is fundamental is related to when (in time) you ask (and perhaps who).

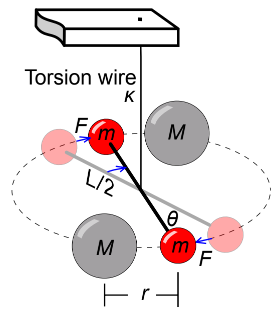

Figure 1: Setup of Cavendish Experiment. Source: wikipedia

So taken today, what do we mean by fundamental physics? What we could start with (this is what I came up with, so please if you disagree let me know), are fundamental constants. Those at least seem to survive the ages. For example, the gravitational constant has been around for many centuries and first determined in 1798 Henry Cavendish.

Cavendish Experiment: This experiment conducted in 1798, perhaps can be considered as the first test of fundamental physics in a modern sense. The goal of this experiment was to determine the value of the gravitational constant G. The experiment set-up is shown in Fig. 1. A dumbbell with two masses m on its ends hangs from a very thin wire. The dumbbell is free to twist. If it twists the wire will provide a restoring torque. Two other heavy masses are then placed with an offset to the dumbbell position without a twist. These two masses produce attractive forces on the dumbbell which will cause it to twist clockwise. After oscillating back and forth with frequency ω, the dumbbell will eventually come at rest at an angle Θ offset compared to the initial position. Using the properties of torque, the properties of a harmonic oscillator, Cavendish knew that he could determine G via

$$G = \frac{4\pi^2 I \Theta r^2}{2Mm\ell T^2}, \tag{1}$$

where $\ell = L/2$ the half-length of the dumbbell, r the distance between the center of each masses m and M, I the moment of inertia of the dumbbell and $T = 2\pi/\omega$. At the time he determined $6.74 \times 10^{-11}\mathrm{m}^3\mathrm{kg}^{-1}\mathrm{s}^{-2}$, which is only 1% away from current estimates. Interestingly, using this value it is also possible to determine the mass of the Earth. That estimate suggests a much higher density then derived from simply measuring the Earth's surface density, which would result in the conclusion that the density deep inside the Earth must be much higher!

Similarly, the speed of light is thought to be a fundamental constant (and first measured by the Michelson experiment in 1881). Can you think of other constants that are considered fundamental?

From Wikipedia: 'The term of "fundamental physical constant" is reserved for physical quantities which, according to the current state of knowledge, are regarded as immutable and as non-derivable from more fundamental principles. ' So at least Wikipedia agrees that the fundamental nature of these quantities depends on the current state of knowledge.

Weinberg, in his book on 'A dream of a final theory' discusses the idea of fundamental. His notion is related to reductionism. In his view, elementary particle physics, as understood today, describes the fundamental basis of everything we see around us. While he argues (and I agree) that it does not mean we are able to compute things on all scales (or make predictions about large interactive macroscopic systems – e.g. consciousness), he does support the idea that there is some fundamental truth, and the rules are set and confined from which all must be derived. Assuming elementary particle physics is the physics that sets these rules, implies this must be fundamental physics.

'Cosmological tests' of fundamental physics would then be to test whether the laws of physics that we think are at the roots of all natural phenomena are true on cosmological scales, or perhaps, to find evidence of the contrary, i.e. to conclude that it can not be reduced to elementary particles. Notably, we are already aware that the current well tested model of particle physics (the standard model) is not the end of the story since it does not play well with gravity. Testing these ideas unfortunately is challenging, requiring higher and higher energies.

Testing fundamental constants using cosmological observations: It is possible to test whether some fundamental constants are truly constant (or abide the notion that the laws of physics are universal) using cosmological observations. The main channel to do this using cosmological observations is by carefully studying the recombination history of our universe. The precise recombination processes, which led to the formation of neutral hydrogen and resulted in the universe becoming almost instantly transparent for CMB photons, depends for example on the mass of the electron m_e and the fine-structure constant α. For a detailed analysis see e.g. `https://arxiv.org/pdf/1406.7482.pdf`.

In Fig. 2 you see what is referred to as the physics cube. It shows the relation between different theoretical corners of physics. It can be argued that the more energy is required to test a theory, the more fundamental it gets. To test a theory of quantum gravity for example, requires us to consider both c, $\hbar$ and G to be non-zero, i.e. we need to test

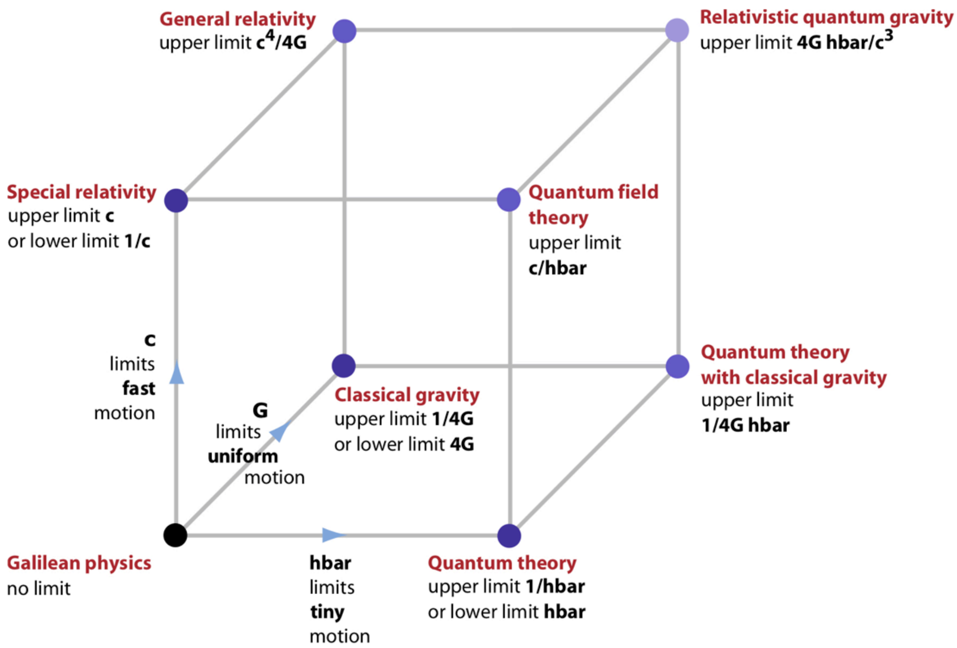

Figure 2: Physics cube (sometimes called Bronshtein's physics cube) showing how 'theories' naturally evolve from one to the other by considering c, $\hbar$ and G. It can be argued that when all of these fundamental constants become relevant at the same time (the relativistic quantum gravity corner of the cube) we are exploring the most fundamental corner of physics. It turns out that we need to a lot of energy (and mass) to probe this regime, making the early universe an ideal testing ground for such theories.

general relativity on very small scales.

The reason why cosmology is suitable to test fundamental physics is that the universe, if we go back far enough in time, was extremely energetic while gravity has played an important role at all times. Studying the cube of fig xx, then suggest we could perhaps explore the corner that concerns the part that is hard to explore otherwise. Obviously we can not go back in time, but the universe as we observe it today, either through the distribution of large scale structures or by observing the relic afterglow of the big bang, the cosmic microwave background (CMB), is a result of the physics in the very early energetic universe. The claim we make, is that by carefully studying these cosmological observations we infer the physics at these very high energies.

This lecture series aims to explain how we go from some theoretical prediction, to an actual (cosmological) observable and make such inference. I plan to discuss our current understanding and future limitations. While there are many different ways (and many great books) to explore this topic, in this lecture series I will focus on cosmological correlation functions that relate to the statistical properties generated by the physical processes in the early universe. While other proposals exist, currently all observations are consistent with the so-called inflationary paradigm, which postulates a brief epoch in the universe, where space rapidly expands (we will get to this in lecture 3). Since inflation happened at such early times, it also occurred when the universe was very energetic ($E \sim 10^{15}$GeV). Within the inflationary paradigm, we can predict statistical properties at the end of inflation, which source the statistical properties of density and radiation fluctuations, which we subsequently observe in the large scale structure and relic background radiation. The amount of information we hope to learn from cosmological information relies critically on the physics of inflation. In particular, from observations we know that inflation has sourced a spectrum[1] of purely scalar fluctuations that is almost scale invariant, adiabatic and Gaussian. If these were all exactly true, then all we can ever hope to learn from cosmological observations is the overall amplitude of (scalar/density) fluctuations. As you can imagine, this severely limits the constraining power of any 'fundamental' physics; by default, every model of the early universe would simply have to make sure it complies with the observed amplitude, and there would be no other way to distinguish between models. Fortunately, data has shown us that the spectrum of fluctuations is not exactly scale invariant. Currently, this only allows to constrain one additional degree of freedom (dof, parameter), but it already provides some useful insight into the physics. Many models of inflation however, predict both the production of tensor fluctuations (gravitational waves) and deviations from Gaussianity[2]. If either of these are observed it would be a monumental discovery and would help further narrow down the number of inflationary scenarios, thereby shedding light on the fundamental physics driving inflation. Note that our current understanding of inflation suggests we are already 'observing' new physics; unless the Higgs is responsible for inflation, inflation must be caused by a degree of freedom not found in the standard model of particle physics.

[1]If you are not familiar with a spectrum, we will discuss this at length. Generally, with a spectrum we mean some function that describes a relation between a scale and an energy or related quantity. Hence scale invariance would imply that this spectrum is constant, and hence there is only a single number, the overall amplitude.

[2]In principle also deviations from adiabaticity, but this will not be the focus on these lectures, please see e.g. [**zz**]

Current effort in observational cosmology try to search for 3 possible routes to learn more about inflation and the early universe:

- Finding deviations from scale invariance beyond a simple one parameter model,
- Detecting relic gravitational waves and,
- Finding traces of primordial non-Gaussianities.

Depending on the pace of the lectures, I hope to discuss all 3 of these. Observationally, these lectures will focus on the cosmic microwave background. Both scale dependence and non-Gaussianity can also be constrained using large scale structures, but in the interest of time, we will limit ourselves to the CMB. Because the CMB was formed early, the primary modes are still relatively small and this allows us to apply linear perturbation theory. By primary I mean those fluctuations (modes) that last scattered or were otherwise unaffected after the CMB was formed. What matters most, in our search for these signals, are foregrounds (e.g. our own galaxy, point sources) and secondaries (e.g. gravitational lensing, SZ effects). As long as we have these under control (besides instrumental noise and systematics), there is principle really no limit to how well you can constrain all 3 of these using the CMB primary modes[3] as long as we increase the resolution of our experiment[4]. Unfortunately, even if we would have a noiseless instrument, we would actually be limited by these two things. For example, detecting primordial gravitational waves requires us to fully qualify and quantify polarized foregrounds from our own galaxy. The search for primordial non-Gaussianities is hampered by classifying all sources of secondaries, in particular on small scales. Note that does not mean that only primordial non-Gaussianities are of cosmological interest. Actually measuring secondaries and using them for cosmological inference is a big part of the industry at the moment and will likely only become more important as observations get more and more precise. For example, (weak) gravitational lensing deflects photons in the CMB which can be observed through to non-Gaussianities in the distribution of temperature and polarization fluctuations (at linear order in the lensing potential it sources a 4-point function, or trispectrum). I will discuss some these secondaries, but mostly in the context of sources of confusion and extra noise when searching for primordial non-Gaussianities.

Note that this lectures series will not be complete and there many more things to learn and to discuss. Material in these lectures can be traced to the following resources, which will actually contain much more information:

- Komatsu: `https://arxiv.org/pdf/1003.6097.pdf`
- Dodelson and Schmidt: Modern Cosmology
- Pajer: `https://www.damtp.cam.ac.uk/user/ep551/cosmology_my_lecture_notes.pdf`
- Baumann: Cosmology and his lectures on Cosmology
- Pimentel's GGI lectures
- Anthony Challinor part III Cosmology: `https://cosmologist.info/teaching/EU/ADC_Structure_formation2.pdf`
- Byrnes: `https://arxiv.org/abs/1411.7002`
- Fergusson and Shellard: `https://arxiv.org/pdf/astro-ph/0612713.pdf`
- Meerburg et al: `https://arxiv.org/pdf/1610.06559.pdf`
- Meerburg et al: `https://arxiv.org/abs/1903.04409`
- Maldacena: `https://arxiv.org/pdf/astro-ph/0210603.pdf`
- Meerburg et al: `https://arxiv.org/pdf/0901.4044.pdf`
- Holman and Tolley: `https://arxiv.org/pdf/0710.1302.pdf`
- Lewis et al: `https://arxiv.org/pdf/1101.2234.pdf`
- Coulton et al: `https://arxiv.org/pdf/2208.12270.pdf`
- Coulton et al: `https://arxiv.org/abs/1912.07619`
- Chen et al: `https://arxiv.org/pdf/hep-th/0605045.pdf`
- Cheung et al: `https://arxiv.org/pdf/0709.0293.pdf`
- Senatore and Zaldarriaga: `https://arxiv.org/abs/1009.2093`
- Pimentel et al: `https://arxiv.org/abs/2203.08128`
- Kalaja et al: `https://arxiv.org/abs/2011.09461`

For a more complete understanding I suggest navigating to these resources. They also contain suggestions and references to more material (yes, it is turtles all the way down). Again, this is not a complete list by any measure!

1.1 Cosmological principle and Correlation functions

In order to do some useful and interesting calculations, we need to first make sure we all have a basic understanding of Cosmology. While I am not sure how much you already know, let us at least derive some of the relevant equations that govern our Universe. This will involve the background evolution. This background evolution is sufficient to remind ourselves why we need something like inflation. The inflationary paradigm solves cosmological conundrums while at the same time it provides a mechanism to produce the seeds of structure. Let us label these quantum seeds as $\delta\phi$, where

[3] at some point there are also corrections due to second order effects in the propagation and sources

[4] On small scales information is washed out by the diffusion of photons, but if were able to push the noise way below the signal, in principle these modes would still provide new information.

the background evolution is driven by $\bar{\phi}(t)$ and we have defined the fluctuations as $\phi(t, \mathbf{x}) = \bar{\phi}(t) + \delta\phi(t, \mathbf{x})$, which is the scalar perturbed part of some unknown field that dominates the very early universe. ϕ is typically computed at the end of inflation, and is a function of space (or momentum) and time. As we will show later, the correlation functions of the fluctuations in ϕ, encode physical properties of the field and possible interactions. Crudely, theoretically what we hope to establish is the following:

$$\delta\phi \to \Psi \to \delta\rho, \Theta. \tag{2}$$

Here we assume the fluctuations sourced in the early universe ($\delta\phi^5$) do not directly couple to matter $\delta\rho$, but do couple to the metric, with metric fluctuations Ψ. Subsequently, gravity will perturb the matter (δ) and radiation fields (Θ). By this simple relation, we then hope by observations of the late-time density fields, to learn something about ϕ. The mapping from left to right could be very complicated. However, by construction the above quantities are small and to lowest order, we can derive linear transfer functions (that preserve the statistical properties of the primordial seeds) that map the initial fluctuations $\delta\phi$ to e.g. the late-time radiation field Θ. Schematically,

$$\Theta \propto \mathcal{F}(\delta\phi), \tag{3}$$

where $\mathcal{F}$ is some functional that typically will involve some integral over time. Note that generally is hard to write down $\mathcal{F}^{-1}$ (basically the transfer functions can be highly oscillatory and at zero you would get infinities).

Our current understanding postulates that the initial fluctuations $\delta\phi$ are described a random (Gaussian[6]) field. The nature of this field is such that at late times we can only infer the statistical properties (the moments of the probability distribution describing the field) of this field, by measuring the statistical properties of late-time observables.

If the initial field is Gaussian, the statistics of the field is entirely described by the first (mean) and second moment (variance), as you could have guessed from the fact that a Gaussian field is described by a Gaussian (which only depends on the mean and the variance).

Since the fields we will consider depend on space (and generally also on time) computing the various moments of the distribution describing the field would depend on many degrees of freedom.

Question 1.1 Consider the n-th moment the probability distribution of this field which has a different value at each position in (3D) space, how does the maximum number of degrees of freedom of the n-th moment depend on n?

Luckily, one key principle and observationally confirmed property is that the universe, on large scales, looks the same for all observers[7] This is referred to as the cosmological principle, which states the universe is both homogeneous and isotropic on large enough scales. If the statistics of the primordial field ϕ is preserved after evolution (by means of a linear transfer), the cosmological principle simplifies the functional form of moments of the observed field (as well as those of the primordial field).

Let us consider a spatial (not necessarily Gaussian) random field $\phi(\mathbf{x})$, i.e. a field ϕ which has a random value at each point in space $\mathbf{x}$ (for example the the field driving inflation). Assume the probability of realising some field configuration is given by $\mathrm{P}[\phi(\mathbf{x})]$, where P is again a functional. We can then write correlation functions as expectation values of products of these random fields at different spatial positions. For example, we can define the 2-point correlation function as

$$\xi(\mathbf{x}, \mathbf{y}) \equiv \langle \phi(\mathbf{x}), \phi(\mathbf{y}) \rangle = \int \mathcal{D}\phi \mathrm{P}[\phi]\phi(\mathbf{x})\phi(\mathbf{y}). \tag{4}$$

Here the integral is a functional integral over field configurations (identical to path integrals in QFT). A similar expression can be written for higher order moments.

The cosmological principle postulates both statistical homogeneity and isotropy. Statistical homogeneity means that if I apply some translation $\mathbf{s}$

$$\mathcal{T}_{\mathbf{s}}\phi(\mathbf{x}) \equiv \phi(\mathbf{x} - \mathbf{s}) \tag{5}$$

the statistical properties of the field should be unchanged, i.e.

$$\mathrm{P}[\phi(\mathbf{x})] = \mathrm{P}[\mathcal{T}_{\mathbf{s}}\phi(\mathbf{x})] \tag{6}$$

For the 2-point correlations function this implies:

$$\xi(\mathbf{x}, \mathbf{y}) = \xi(\mathbf{x} - \mathbf{x}, \mathbf{y} - \mathbf{s}). \tag{7}$$

The only way for this to be true is that the 2-point function can be a function of the vector separation between the two positions only, i.e. $\xi(\mathbf{x}, \mathbf{y}) = \xi(\mathbf{x} - \mathbf{y})$.

Statistical isotropy implies that the statistical properties of the field are invariant under rotations:

$$\mathcal{R}\phi(\mathbf{x}) \equiv \phi(\mathbf{R}^{-1}\mathbf{x}), \tag{8}$$

where $\mathbf{R}$ is a rotation matrix. Again, applying this for the 2-point correlation function, invariance implies:

$$\xi(\mathbf{x}, \mathbf{y}) = \xi(\mathbf{R}^{-1}\mathbf{x}, \mathbf{R}^{-1}\mathbf{y}). \tag{9}$$

[5] We will later introduce the field ϕ that sources the initial conditions. For practical purposes, we will also who that it will be more convenient to work directly with curvature fluctuations $\zeta, \mathcal{R}$ as apposed to $\delta\phi$.

[6] Or not!

[7] note that we obviously can not place an observer everywhere, however if the the universe did not have this property, computations would lead to very different results which do not match current observations.

Combining this with statistical homogeneity we get:

$$\xi(\mathbf{x},\mathbf{y}) = \xi(\mathbf{R}^{-1}(\mathbf{x}-\mathbf{y})) \rightarrow \xi(\mathbf{x},\mathbf{y}) = \xi(|\mathbf{x}-\mathbf{y}|), \tag{10}$$

i.e. the 2-point correlator can only depend on the distance modulus.

Throughout this reader we will often work in Fourier space. We can repeat the above in Fourier space:

$$\phi_{\mathbf{k}} = \int d^3\mathbf{x}\phi(\mathbf{x})\mathbf{e}^{-\mathbf{i}\mathbf{k}\cdot\mathbf{x}}, \tag{11}$$

and the reverse:

$$\phi(\mathbf{x}) = \int \frac{d^3\mathbf{k}}{(2\pi)^3}\phi_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{x}} \tag{12}$$

I we apply we translation to the field in momentum space we pick up a phase factor:

$$\begin{aligned} \mathcal{T}_{\mathbf{s}}\phi_{\mathbf{k}} &= \int d^3\mathbf{x}\phi(\mathbf{x}-\mathbf{s})e^{-i\mathbf{k}\cdot\mathbf{x}} \\ &= \int d^3\mathbf{x}'\phi(\mathbf{x}')\mathbf{e}^{-\mathbf{i}\mathbf{k}\cdot\mathbf{x}'}\mathbf{e}^{-\mathbf{i}\mathbf{k}\cdot\mathbf{s}}, \end{aligned} \tag{13}$$

where we simply replace integration variable $\mathbf{x} \rightarrow \mathbf{x}' = \mathbf{x} - \mathbf{s}$.

Demanding invariance of the 2-point correlator (or power spectrum) in Fourier space:

$$\langle \phi_{\mathbf{k}}\phi^*_{\mathbf{k}'}\rangle = \langle \phi_{\mathbf{k}}\phi^*_{\mathbf{k}'}\rangle e^{-i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{s}}, \tag{14}$$

which implies

$$\langle \phi_{\mathbf{k}}\phi^*_{\mathbf{k}'}\rangle = F(\mathbf{k})\delta(\mathbf{k}-\mathbf{k}'). \tag{15}$$

$F(\mathbf{k})$ is some function. The total momentum is thus conserved, or different momenta are uncorrelated.

Similarly, we can apply rotations

$$\begin{aligned} \mathcal{R}\phi_{\mathbf{k}} &= \int d^3\mathbf{x}\phi(\mathbf{R}^{-1}\mathbf{x})e^{-i\mathbf{k}\cdot\mathbf{x}} \\ &= \int d^3\mathbf{x}\phi(\mathbf{R}^{-1}\mathbf{x})e^{-i\mathbf{R}^{-1}\mathbf{k}\cdot\mathbf{R}^{-1}\mathbf{x}} \\ &= \phi_{\mathbf{R}^{-1}\mathbf{k}} \end{aligned} \tag{16}$$

Note that in going from the first line to the second we use that fact that we rotate two vectors with the same matrix, their dot product should remain invariant.

If we now combine translational and rotational invariance like we did in position space:

$$\langle \mathcal{R}\phi_{\mathbf{k}}[\mathcal{R}\phi_{\mathbf{k}'}]^*\rangle = \langle \phi_{\mathbf{R}^{-1}\mathbf{k}}\phi^*_{\mathbf{R}^{-1}\mathbf{k}'}\rangle = F(\mathbf{R}^{-1}\mathbf{k})\delta(\mathbf{R}^{-1}(\mathbf{k}-\mathbf{k}')) = F(\mathbf{k})\delta(\mathbf{k}-\mathbf{k}'). \tag{17}$$

Since $\delta(\mathbf{R}^{-1}(\mathbf{k}-\mathbf{k}')) = \det\mathbf{R}\delta(\mathbf{k}-\mathbf{k}') = \delta(\mathbf{k}-\mathbf{k}')$, because $\det\mathbf{R} = 1$, this equality implies $F(\mathbf{k}) = F(\mathbf{R}^{-1}\mathbf{k})$ or $F(\mathbf{k}) = F(k)$. This allows us to define the power spectrum $\Delta^2_\phi(k)$ of a homogeneous and isotropic field $\phi_{\mathbf{k}}$ by

$$\langle \phi_{\mathbf{k}}\phi^*_{\mathbf{k}'}\rangle = \frac{2\pi^2}{k^3}(2\pi)^3\delta(\mathbf{k}-\mathbf{k}')\Delta^2_\phi(k). \tag{18}$$

The normalisation $2\pi^2/k^3$ makes sure that the power spectrum $\Delta^2_\phi(k)$ is dimensionless if $\phi_{\mathbf{k}}$ is.

Remark Note that the above is a scalar, i.e. the final correlation function depends on a single degree of freedom. Generally, it could have depended on 6 (3 for both $\mathbf{x}$ and $\mathbf{y}$). That it does not is because translational and rotational invariance apply constraints. In Fourier space translational invariance implies conservation of momentum, reducing the total number of degrees of freedom by 3. Rotational invariance allows one to pick a 2-dimensional plane defined by the remaining momenta and adjust the axes such that one of the momenta is along one of the axes of the plane. For a single momentum, as would be the case for the 2-point correlators (or power spectrum) this leads to a reduction of another 2 degrees of freedom, i.e. $6-(3+2) = 1$. While for example for the 3-point correlation function, or bispectrum in Fourier space, it is simply the 3 possible rotations. Generally, for a translational and rotational invariant n-point correlator, with $n > 2$, we expect $3(n-2)$ degrees of freedom.

Question 1.2 Please check the above for a 3- and 4-point correlation function.

In position space the the correlation function is the Fourier transform of the power spectrum, i.e.

$$\begin{aligned} \langle \phi(\mathbf{x})\phi(\mathbf{y})\rangle &= \int\int \frac{d^3\mathbf{k}}{(2\pi)^{3/2}}\frac{d^3\mathbf{k}'}{(2\pi)^{3/2}}\langle \phi_{\mathbf{k}}\phi^*_{\mathbf{k}'}\rangle e^{i\mathbf{k}\cdot\mathbf{x}}e^{-i\mathbf{k}\cdot\mathbf{y}} \\ &= \int \frac{d^3\mathbf{k}}{(2\pi)^{3/2}}\int \frac{d^3\mathbf{k}'}{(2\pi)^{3/2}}\frac{2\pi^2}{k^3}\delta(\mathbf{k}-\mathbf{k}')\Delta^2_\phi(k)e^{i\mathbf{k}\cdot\mathbf{x}}e^{-i\mathbf{k}\cdot\mathbf{y}} \\ &= \frac{1}{4\pi}\int \frac{d^3\mathbf{k}}{k^3}\Delta^2_\phi(k)e^{-i\mathbf{k}\cdot(\mathbf{x}-\mathbf{y})} \\ &= \frac{1}{4\pi}\int \frac{dk}{k}\Delta^2_\phi(k)\int d\Omega e^{-i\mathbf{k}\cdot(\mathbf{x}-\mathbf{y})}. \end{aligned} \tag{19}$$

The angular integral can be performed if we assume $\mathbf{x}-\mathbf{y}$ to be along the z axis in Fourier space. Then we can write

$\mathbf{k}\cdot(\mathbf{x}-\mathbf{y}) = k|\mathbf{x}-\mathbf{y}|\cos\Theta$, with Θ the angle between the $\mathbf{k}$ and the $\hat{z}$ direction. Setting $\mu = \cos\Theta$ we can write

$$\begin{aligned}\int d\Omega e^{-i\mathbf{k}\cdot(\mathbf{x}-\mathbf{y})} &= \int_0^{2\pi}\int_0^{\pi} d\phi \sin\Theta d\Theta e^{-i\mathbf{k}\cdot(\mathbf{x}-\mathbf{y})} \\ &= 2\pi\int_{-1}^{1} d\mu e^{ik|\mathbf{x}-\mathbf{y}|\mu} \\ &= 4\pi j_0(k|\mathbf{x}-\mathbf{y}|), \end{aligned} \tag{20}$$

where $j_0(x) = \sin x/x$ is the Spherical Bessel functoin of order zero. We thus find

$$\xi(\mathbf{x},\mathbf{y}) = \int \frac{dk}{k}\Delta^2_\phi(k) j_0(k|\mathbf{x}-\mathbf{y}|), \tag{21}$$

which does only depend on the distance modules of the vectors $\mathbf{x}$ and $\mathbf{y}$ as we derived earlier.

Both the position and momentum space correlation function and power spectra respectively live in 4D Minkowski space. The CMB is the last scattering surface of photons in the early universe. Observationally, it lives on the surface of a sphere with a radius the size of the current universe. The information content in the CMB is compressed to this surface. To express the fact that the CMB lives on this surface we do not work in full 3D position or Fourier space, but instead we express our field in terms of spherical harmonics:

$$\Theta(\hat{\mathbf{n}}) = \sum_{\ell m} a_{\ell m} Y_{\ell m}(\hat{\mathbf{n}}), \tag{22}$$

where Θ is measured on some position in the sky $\hat{\mathbf{n}}$ and $a_{\ell m}$ and $Y_{\ell m}(\hat{\mathbf{n}})$ are the spherical harmonic coefficients and functions respectively. The above expression holds for a scalar observable, such as the Temperature of the CMB. Note that $\hat{\mathbf{n}}$ depends on two coordinates, e.g. ψ and θ; we do not have a radial coordinate since the CMB lives on the surface of the sphere.

We can perform similar exercises as before. Instead of working with the Fourier modes $\phi_{\mathbf{k}}$ we work with the spherical harmonic coefficients $a_{\ell m}$. As before, statistical isotropy implies that statistics are invariant under rotations. We define the rotation in spherical harmonic space as

$$\mathcal{R}Y_{\ell m}(\hat{\mathbf{n}}) = \sum_{m'=-\ell}^{\ell} \mathbf{R}^{(\ell)}_{mm'} Y_{\ell m'}. \tag{23}$$

For the **angular** power spectrum to be invariant under rotations then implies

$$\langle a_{\ell_1 m_1} a^*_{\ell_2 m_2}\rangle = \sum_{m_1' m_2'} \langle a_{\ell_1 m_1'} a^*_{\ell_2 m_2'}\rangle \mathbf{R}^{(\ell_1)}_{m_1 m_1'} \mathbf{R}^{(\ell_2)*}_{m_2 m_2'}. \tag{24}$$

The above is only true if the power spectrum is diagonal, i.e.

$$\langle a_{\ell_1 m_1'} a^*_{\ell_2 m_2'}\rangle = C_{\ell_1}\delta_{\ell_1\ell_2}\delta_{m_1' m_2'}. \tag{25}$$

In that case

$$\begin{aligned}\langle a_{\ell_1 m_1} a^*_{\ell_2 m_2}\rangle &= C_{\ell_1}\delta_{\ell_1\ell_2}\sum_{m_1'} \mathbf{R}^{(\ell_1)}_{m_1 m_1'} \mathbf{R}^{(\ell_2)*}_{m_2 m_1'} \\ &= C_{\ell_1}\delta_{\ell_1\ell_2}\delta_{m_1 m_2}. \end{aligned} \tag{26}$$

where we used that the rotation matrices are orthonormal. Similar expressions can be derived for the angular bispectrum and trispectrum which we will leave as an exercise. Note that we only had to invoke statistical isotropy, and not statistical homogeneity.

Similar to the Fourier power spectrum, we can relate the angular power spectrum to the correlation function in position space as

$$\begin{aligned}\langle\Theta(\hat{\mathbf{n}}_1)\Theta(\hat{\mathbf{n}}_2)\rangle &= \sum_{\ell_1 m_1 \ell_2 m_2} \langle a_{\ell_1 m_1} a^*_{\ell_2 m_2}\rangle Y_{\ell_1 m_1}(\hat{\mathbf{n}}_1) Y^*_{\ell_2 m_2}(\hat{\mathbf{n}}_2) \\ &= \sum_{\ell_1} C_{\ell_1}\sum_{m_1} Y_{\ell_1 m_1}(\hat{n}_1) Y^*_{\ell_1 m_1}(\hat{n}_2) \\ &= \frac{1}{4\pi}\sum_{\ell}(2\ell+1) C_\ell P_\ell(\cos\theta_{1,2}), \end{aligned} \tag{27}$$

where $\theta_{1,2}$ is the angle between $\hat{\mathbf{n}}_1$ and $\hat{\mathbf{n}}_2$, P is a Legendre Polynomial and we took the liberty to change to variable ℓ on the last line. Note that one of the reasons why it is preferred to work in spherical harmonic space over position space, is evident by the dependence on $\cos\theta_{1,2}$; in the correlation function, different angles are correlated. In the angular power spectrum, under the assumption of statistical isotropy, different multiples ℓ are uncorrelated (similar to how different $\mathbf{k}$ are uncorrelated in the Fourier power spectrum).

Coming back to Eq.(2), while we can not measure a single realization of $\delta\phi$, we instead measure correlation (or equivalently $n-1$ spectra) of observables. These are theoretically related to the primordial spectra, after projecting these primordial spectra forward with some linear transfer function. By minimizing the variance between an observed and a theoretical spectrum, we infer cosmological parameters hidden in the transfer functions as well as the details of $\Delta^2_\phi(k)$. Ideally, we would like to do the same for higher order correlation functions, but this currently not computationally feasible. I will later discuss how to forecast parameters constraints using Fisher forecasts (specifically primordial non-Gaussianity).

1.2 An expanding Universe

In order to make the mapping from the early universe to a cosmological observable requires knowing the mappings $\mathcal{F}$, or (linear) transfer functions. To obtain this transfer function requires deriving how fluctuations evolve as they are imprinted at the end of inflation, to today. This requires us to qualify how different components in the universe interact (electromagnetically and gravitationally). As a first step we derive the background equations of motion of the fields (matter, radiation, etc). In order to that we need to specify the metric of the Universe. Many of the resources provided in the introduction will derive the metric of the universe, here I will simply tell you that it can be described by a Friedmann-Lemaitre-Robertson-Walker (FLRW) metric (with $k = 0$, i.e. no curvature)

$$ds^2 = a(t)^2 \left[-d\tau^2 + \delta_{ij} dx^i dx^j\right]. \tag{28}$$

Here $a(t)$ is the scale factor of the universe and τ is conformal time: $a d\tau = dt$[8] t is sometimes referred to cosmological time (or time). Note that x are comoving spatial coordinates:

$$\Delta x_{\rm phys} = \sqrt{a^2 \delta_{ij} \Delta x^i \Delta x^j} = a|\Delta x| \tag{29}$$

i.e.they are related to physical spatial intervals by a factor of a. We define the Hubble parameter as $H(t) = a'/a$, where $'$ represents derivatives w.r.t. to cosmological time. Similarly we can define $\mathcal{H} = \dot{a}/a$ where now derivatives are w.r.t. conformal time τ. The Hubble parameter literally measures the change in the scale factor over time, in units of the scale factor. The scale factor is a relative measure (and dimensionless), and the Hubble parameter tells us the rate of change of the relative scales in the universe. A positive Hubble parameter tells us that the universe is growing and a negative value would suggest it is shrinking. Note however that because it concerns the relative scale, it can still be true that the universe is infinite in size (physically). Imagining the universe at some initial time on a fixed coordinate grid, a positive Hubble parameter would simply tell us that the distance between grid points marking that grid grows over time.

The Christoffel symbols of any metric in General Relativity can be derived from

$$\Gamma^\mu_{\alpha\beta} = \frac{1}{2} g^{\mu\gamma}(g_{\alpha\gamma,\beta} + g_{\beta\gamma,\alpha} - g_{\alpha\beta,\gamma}), \tag{30}$$

where $_,$ denote derivatives.

Question 1.3 Show that the only non-vanishing Christoffel symbols for the (unperturbed) flat FRLW metric are:

$$\Gamma^0_{ij} = Ha^2\delta_{ij}, \quad \Gamma^j_{i0} = \Gamma^i_{j0} = H\delta_{ij}. \tag{31}$$

Note that we have set $c = 1$ throughout.

The Einstein equations prescribe the relation between the geometry (described the metric) and energy (described by the energy momentum tensor):

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} = 8\pi G T_{\mu\nu}. \tag{32}$$

We need to unpack a couple of things here. $R_{\mu\nu}$ is the Ricci tensor and $R = g^{\mu\nu} R_{\mu\nu}$ the Ricci scalar. Both are fully determined by the geometry of spacetime, specifically through the Christoffel symbols:

$$R_{\mu\nu} = \Gamma^\lambda_{\mu\nu,\lambda} - \Gamma^\lambda_{\mu\lambda,\nu} + \Gamma^\lambda_{\lambda\rho}\Gamma^\rho_{\mu\nu} - \Gamma^\rho_{\mu\lambda}\Gamma^\lambda_{\nu\rho}. \tag{33}$$

Question 1.4 Show that the only two non-vanishing components of the Ricci tensor are given by

$$R_{00} = -3\frac{a''}{a}, \tag{34}$$

$$R_{ij} = \left[-\frac{1}{3} R_{00} + 2H^2\right] a^2 \delta_{ij}. \tag{35}$$

and that

$$R = 2\left[-R_{00} + 3H^2\right]. \tag{36}$$

From here derive that

$$G_{00} = -3H^2 \tag{37}$$

$$G_{ij} = \left[\frac{2}{3} R_{00} - H^2\right] a^2 \delta_{ij}. \tag{38}$$

(please check)

The energy-momentum tensor for a perfect fluid is given by:

$$T_{\mu\nu} = (\rho + P) U_\mu U_\nu + P g_{\mu\nu}. \tag{39}$$

Here ρ is the energy density, P the pressure, both in the rest frame of the fluid. U^μ is the four-velocity:

$$U^\mu \equiv \frac{dx^\mu}{d\tau}. \tag{40}$$

For a comoving observer ("expanding with a") at rest $U^\mu = (1, 0, 0, 0)$. Conservation of energy-momentum can be written

[8] The FLRW metric is conformally equivalent to Minkowski, i.e. what is in brackets looks like Minkowski and a only depends on time.

as

$$\nabla_\mu T^\mu{}_\nu = 0, \tag{41}$$

where ∇_μ is the covariant derivative:

$$\nabla_\mu T^\mu{}_\nu = \partial_\mu T^\mu{}_\nu + \Gamma^\mu_{\mu\lambda} T^\lambda{}_\nu - \Gamma^\lambda_{\mu\nu} T^\mu{}_\lambda \tag{42}$$

Question 1.5 Using the relations derived above for the Christoffel symbols in a homogeneous and isotropic universe, show that conservation of the stress-energy tensor results in the following relations between the pressure and the density:

$$\rho' + 3H(\rho + P) = 0, \tag{43}$$

considering the $\nu = 0$ component of Eq. (42). (please check)

The equation above describes energy conservation in an expanding universe. Since time translations are broken it is different from energy conservation in Euclidean space.

We can parametrize the usual components in the universe in terms of an equation of state, which relates the pressure to the energy density, i.e.

$$P/\rho = \omega. \tag{44}$$

Using energy conservation this results in the following general relation for the energy density of a component in an expanding universe:

$$\rho \propto a^{-3(1+\omega)}. \tag{45}$$

For radiation (for neutrinos before they become non-relativistic) $\omega_{\rm r} = 1/3$ for matter (dark and ordinary or baryonic) $\omega_{\rm m} = 0$ and for Dark Eenery $\omega_\Lambda = -1$.

Question 1.6 Using the Einstein field equations (Eq. (32)), show that the time-time component of these equations leads to

$$H^2 = \frac{8\pi G}{3}\rho, \tag{46}$$

which is generally known as the Friedmann equation.
Next show that the spatial part of the Einstein field equations results in the second Friedmann equation:

$$\frac{a''}{a} = -\frac{4\pi G}{3}(\rho + 3P). \tag{47}$$

In cosmology, we typically define a critical density, which is the energy density of the universe today, if there was no spatial curvature[9]:

$$\rho_{\rm crit} \equiv \frac{3H_0^2}{8\pi G} = 1.9 \times 10^{-29} h^2 {\rm g\ cm}^{-3}, \tag{49}$$

where we define $H_0 = H(t=0) = 100h{\rm km/Mpc/s}$ (and the measured value of $h = \mathcal{O}(1)$). The above density corresponds to about 10^{-5} protons per cubic cm. With the definition of the critical density, all other densities are conveniently rewritten as

$$\Omega_i = \frac{\rho_i}{\rho_{\rm crit}}, \tag{50}$$

known as the dimensionless density parameter, which allows us to write

$$H^2 = H_0^2\left(\Omega_{\rm r} a^{-4} + \Omega_{\rm m} a^{-3} + \Omega_\Lambda\right), \tag{51}$$

i.e. the sum of the radiation (r), matter (m) and dark energy (Λ) densities respectively. For a critical universe, i.e. no curvature, the sum of these should add up to 1 throughout the history of the universe. Note again that our normalisation is set that today $a = 1$.

Question 1.7 Derive the evolution of the scale factor as a function of time assuming a radiation dominated universe, a matter dominated universe and a universe dominated by a cosmological constant.
Next, use any (computer) language of choice to solve the scale factor as a function of time for a 3 components universe, where $\Omega_r = 0.00005$, $\Omega_m = 0.29995$ and $\Omega_\Lambda = 0.7$. Show that your analytically derived scalings match inside the regimes where each component dominates. When do we expect Λ to start dominating the universe?

[9] For sake of simplicity we have assumed no curvature throughout, but any modern textbook on cosmology will show you that the the Friedmann equation is altered as

$$H^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2}. \tag{48}$$

Here k is the spatial curvature of the universe. Interestingly note that the 'energy density' of this term contributes negatively to the expansion and has $\omega = -1/3$.

1.3 Photons in an expanding universe

For a massless particle

$$P_\mu P^\mu = m^2 = 0, \tag{52}$$

i.e. the magnitude of the 4-momentum is zero. Massless particles, such as photons, propagate through the universe (after they decouple from the baryon plasma) following a geodesic, where the geodesic is given by

$$\frac{dP^\mu}{d\lambda} = -\Gamma^\mu_{\nu\rho} P^\nu P^\rho. \tag{53}$$

which describes the trajectory of a particle in the absence of forces (the equivalent of $dx^2/dt^2 = 0$ in classical dynamics). The affine parameter is defined such that $P^\mu = dx^\mu/d\lambda$. The (unperturbed) photon 4-momentum is given by $P^\mu = (E, E\hat{p}^i)$. We can write:

$$\frac{d}{d\lambda} = \frac{dx^0}{d\lambda}\frac{d}{dx^0} = E\frac{d}{dt}. \tag{54}$$

The zeroth component of the geodesic equation (53) then becomes:

$$E\frac{dE}{dt} = \Gamma^0_{ij} P^i P^j = \delta_{ij} H a^2 P^i P^j. \tag{55}$$

Using the fact that the photon is massless and the magnitude of its 4-momentum vanishes

$$g_{\mu\nu} P^\mu P^\nu = -E^2 + \delta_{ij} a^2 P^i P^j = 0 \tag{56}$$

we can rewrite the photon geodesic equation as

$$\frac{1}{E}\frac{dE}{dt} = -H \tag{57}$$

Question 1.8 What is a general solution this equation? Does this solution make sense?

After we have derived the background evolution, can we perturb both the metric and the components in the universe, to derive our linear transfer functions.

Lecture 2: Perturbations

2.4 Scalar-vector-tensor decomposition

To compute n-point correlation functions (or $n-s$-spectra) fluctuations in observable quantities, requires us to perturb the Einstein equations (32). These equations, which relate the Einstein tensor to the energy-momentum tensor, can be perturbed by scalar, vector or tensor perturbations. Computing how those perturbations evolve in an expanding universe would become very complicated if initial perturbations in each of these would source new fluctuations in the others. The scalar-vector-tensor (SVT) decomposition theorem states that at linear order these perturbations evolve independently and will not mix.

The most general perturbed spacetime can be written as

$$ds^2 = a^2\left[-(1+2A)d\tau^2 - B_i dx^i d\tau + (\delta_{ij} + 2E_{ij} dx^i dx^j)\right]. \tag{58}$$

Here A, B_i and E_{ij} are functions of both space and time. The factors of 2 and the labelling of the functions here are quite arbitrary, but fairly standard.

Let us now decompose these perturbations into pure scalar, pure vector and pure tensor perturbations. First, the three vector B_i can be split into the gradient of a scalar B and a divergenceless vector $\hat{B}_i$, i.e. $\partial^i \hat{B}_i = 0$, where the $\hat{}$ denotes that the object is divergenceless:

$$B_i = B_{,i} + \hat{B}_i \tag{59}$$

E_{ij} is a rank 2 symmetric tensor (we assume we remain in an isotropic universe), which can be split in a scalar, and a divergenceless vector and tensor contribution:

$$E_{ij} = C\delta_{ij} + \left(\partial_i\partial_j - \frac{1}{3}\delta_{ij}\nabla^2\right)E + \frac{1}{2}\left(\partial_i\hat{E}_j + \partial_j\hat{E}_i\right) + \hat{E}_{ij}. \tag{60}$$

Note that these are all contributions we can can write using just tensor, scalars and vectors and derivative's thereoff. The first contribution on the RHS contains the trace, i.e. $E^i{}_i = C$, while all the other terms are traceless. After SVT decomposition, we end up with 10 degrees of freedom; 4 scalar functions (A, B, C, E), 2 divergenceless vectors $\hat{B}_i$ and $\hat{C}_i$ and 1 traceless symmetric tensor $\hat{E}_{ij}$ (which is another 4 degrees of freedom). To reiterate the main point, at linear order the evolution of these variables as described by the Einstein equations do not mix, i.e. they evolve independently.

For more discussion on this see Dodelson Chapter 5, sec 5.4. Baumann Chapter 6. section 6.1, and Enrico Pajer's lecture notes Chapter 10.

2.5 Perturbations

Given the scalar-vector-decomposition, their perturbations do not mix and we can independently derive the evolution of scalar, vector and tensor degrees of freedom in our universe. For simplicity, let us consider only scalar perturbations since those will have the largest effect on the observed fluctuations in our universe[10] We write

$$
\begin{aligned}
g_{\mu\nu}(t,\mathbf{x}) &= \bar{g}_{\mu\nu}(t) + h_{\mu\nu}(t,\mathbf{x}), && (61)\\
T_{\mu\nu}(t,\mathbf{x}) &= \bar{T}_{\mu\nu}(t) + \delta T_{\mu\nu}(t,\mathbf{x}) && (62)
\end{aligned}
$$

Let us start by perturbing the metric, where $\bar{g}_{\mu\nu}(t) = \mathrm{diag}(-1, a^2, a^2, a^2)$ and we define the components of $h_{\mu\nu}(t,\mathbf{x})$ using the SVT decomposition above

$$
\begin{aligned}
h_{00} &= -2A\\
h_{0i} &= -aB_{,i}\\
h_{ij} &= a^2(2\delta_{ij}\psi - 2E_{,ij}). && (63)
\end{aligned}
$$

As before, we have two scalars in the spatial part of the metric because we want to able to perturb on and off diagonal components separately. We have identified $C = \psi$ and $E = -E$ compared to the expansion in Eq. (58) (to match to Dodelson's notation).

The spacetime interval is unchanged when we change coordinates, i.e.

$$\tilde{g}_{\alpha\beta}(\tilde{t},\tilde{\mathbf{x}})d\tilde{x}^\alpha d\tilde{x}^\beta = g_{\mu\nu}(t,\mathbf{x})dx^\mu dx^\nu. \qquad (64)$$

Hence

$$\tilde{g}_{\alpha\beta}(\tilde{t},\tilde{\mathbf{x}})\frac{d\tilde{x}^\alpha}{dx^\mu}\frac{d\tilde{x}^\beta}{dx^\nu} = g_{\mu\nu}(t,\mathbf{x}). \qquad (65)$$

Next we consider a coordinate change

$$x^\mu \to \tilde{x}^\mu = x^\mu + \epsilon^\mu(t,\mathbf{x}) \qquad (66)$$

where we define $\epsilon^\mu = (\epsilon^0, \delta^{ij}\epsilon^S_{,j})$, where S stands for scalar, since in principle we could allow an additional vector component. Applying such a general transformation allows us to deduce how we should change the scalar functions to keep the perturbed spacetime interval invariant. Note that this is not the same as deriving how the scalar functions change themselves under a coordinate transformation.

As an example, let us consider A. We set $\mu = \nu = 0$. On the left-hand side we have to sum over all indices α and β. However, when $\alpha = 0$ and $\beta = j$, $\tilde{g}_{0j} = -a\tilde{B}_{,j}$. This is a first order perturbation. However $\partial\tilde{x}^i/\partial t$ is proportional to ϵ^S, which is a first-order variable. Hence this contribution will be second order. A similar argument can be made for $\alpha = i$ and $\beta = j$. Using this, we are only left with the terms associated with $\alpha = \beta = 0$, which lead to

$$
\begin{aligned}
-(1+2\tilde{A})\left(\frac{\partial\tilde{t}}{\partial t}\right) &= -(1+2\tilde{A})\left(1+\frac{\partial\epsilon^0}{\partial t}\right)^2\\
&\simeq -1 - 2\tilde{A} - 2\frac{\partial\epsilon^0}{\partial t}. && (67)
\end{aligned}
$$

Equating this to the RHS of Eq. (65) yields

$$A \to \tilde{A} = A - \frac{1}{a}\dot{\epsilon}^0, \qquad (68)$$

where we moved to conformal time coordinates.

Question 2.9 Show that

$$
\begin{aligned}
\tilde{\psi} &= \psi - H\epsilon^0,\\
\tilde{B} &= B + \dot{\epsilon}^S - \frac{\epsilon^0}{a},\\
\tilde{E} &= E + \epsilon^S. && (69)
\end{aligned}
$$

Effectively then there are 4 scalar functions (A, B, E, ψ), and 2 coordinate transformations (ϵ^0, ϵ^S). Hence we can always choose our coordinate transformations such that in that frame 2 functions are 0. By doing this, we fix the gauge, i.e. we choose coordinates that correspond to the constant hypersurfaces of some of the perturbations, so that those perturbations appear constant.

It is possible to define variables (scalar functions) that are gauge independent, i.e variables that do not change under a coordinate transformation. These were first identified by Bardeen:

$$
\begin{aligned}
\Phi_A &\equiv A + \frac{1}{a}\frac{\partial}{\partial\tau}\left[a(\dot{E} - B)\right]\\
\Phi_B &\equiv -\psi + aH(B - \dot{E}). && (70)
\end{aligned}
$$

The usefullness of these variables is that in cosmology we tend to fix a gauge that is particularly convenient for the problem we are considering. Suppose that we now want to change to another gauge after we computed some physical quantities. We can then use these gauge independent variables to relate the quantities in these gauges. Note that *observable* quantities are always gauge independent.

Note that in order to perform computations that involve the Einstein equations, requires us to also perturb the stress-energy tensor. We would follow the same procedure as above, i.e. we would demand the perturbed energy-momentum

[10] Here we will not be concerned with vector perturbations, because they decay in an FRLW universe. Radiation couples to both tensor and scalar perturbations in the metric, which allows us to detect/constrain gravitational waves sources in the early universe as will be discussed in Lecture 4.

tensor to remain invariant under coordinate transformations. For pure scalar perturbations we define our perturbed energy-momentum tensor via (for details see Baumann Chapter 6, sec 6.1.2 and Enrico's lecture notes Chapter 10).

$$\begin{aligned}\delta T_{00} &= \delta\rho \\ \delta T_{0j} &= -(P+\rho)v_{,i} \\ \delta T_{ij} &= a^2(\delta_{ij}\delta p + \pi^S_{,ij}),\end{aligned} \tag{71}$$

Again we have 4 scalar functions ($\delta\rho, \delta p, v$ and π^S). Note that there are no off-diagonal components along the temporal part in the unperturbed metric. One can derive (see page 213 of Baumann) the perturbed lower index fluid velocity $U_\mu = a(-1-A, v_i + B_i)$, with A the temporal scalar metric perturbation and B_i the vector perturbations as defined in Eq. (59). v_i is the bulk velocity which can be decomposed similar to Eq. (59):

$$v_i = v_{,i} + \hat{v}_i. \tag{72}$$

If there are only scalar perturbations, $v_i = v_{,i}$ and $B_i = 0$. π^S is the anisotropic inertia which is a property of a given fluid that needs to be specified to close the system of equations. For a perfect fluid $\pi^S = 0$.

Remark –Some Gauge choices–

Conformal Newtonian Gauge: If we set $B = E = 0$, we find

$$ds^2 = a^2\left[-(1+2\Psi)d\tau^2 + (1-2\Phi)\delta_{ij}dx^i dx^j\right]. \tag{73}$$

after identifying $A = \Psi$ and $\psi = -\Phi$. This is the conformal Newtonian Gauge (which is sometimes called the longitudinal gauge).

Synchronous Gauge: If we set $A = B = 0$, we find

$$ds^2 = dt^2 + a^2(\delta_{ij}(1+2\psi) - 2E_{,ij})dx^i dx^j. \tag{74}$$

In this gauge clocks for observers tick at the same rate, hence the name Synchronous.

Spatially Flat Gauge: If we set $\psi = E = 0$, we find

$$ds^2 = -(1+2A)dt^2 - 2aB_{,i}dx^i dt + a^2 dx^i dx^j. \tag{75}$$

Evidently, in this gauge the spatial part is flat.

Comoving Gauge: We set $v = B = 0$.

Two other common gauge invariant quantities are

$$\zeta = -\psi - \frac{1}{3}\nabla^2 E + \mathcal{H}\frac{\delta\rho}{\dot\rho} \tag{76}$$

$$\mathcal{R} = -\psi - \frac{1}{3}\nabla^2 E - \mathcal{H}(v+B) \tag{77}$$

These are called curvature perturbations because they reduce to the intrinsic curvature of the spatial slices in the uniform density ($\delta\rho = 0$) and comoving gauges respectively. For constant time, the general expression for the perturbed metric is given by:

$$g_{ij} \equiv \bar{g}_{ij} + h_{ij} = a^2((1+2\psi)\delta_{ij} - 2E_{,ij}). \tag{78}$$

We can compute the 3-curvature using $R_{(3)} = g_{ij}R^{ij}_{(3)}$ and

$$R^{(3)}_{ij} = \Gamma^k_{ij,k} - \Gamma^k_{jk,j} + \Gamma^k_{kl}\Gamma^l_{ij} - \Gamma^l_{ik}\Gamma^k_{jl}. \tag{79}$$

We find (check!)

$$a^2 R_{(3)}[g_{ij}] = -4\nabla^2\psi - 2\partial_i\partial_j E^{ij} \tag{80}$$

$$= 4\nabla^2\left(-\psi - \frac{1}{3}\nabla^2 E\right). \tag{81}$$

This is the reason why these two components are referred to as curvature perturbations. On super (Hubble) horizon scales, i.e. $k \ll \mathcal{H}$, $\zeta \to \mathcal{R}$. For so-called adiabatic perturbations (only one relevant degree of freedom during inflation), on these scales $\zeta, \mathcal{R}$ are constant.

Question 2.10 Using the Newtonian Gauge show that the perturbed Christoffel symbols are given by

$$\Gamma^0_{00} = \mathcal{H} + \dot\Psi \tag{82}$$

$$\Gamma^i_{00} = \partial^i\Psi \tag{83}$$

$$\Gamma^0_{i0} = \partial_i\Psi \tag{84}$$

$$\Gamma^0_{ij} = \mathcal{H}\delta_{ij} - \left[\dot\Phi + 2\mathcal{H}(\Phi+\Psi)\right] \tag{85}$$

$$\Gamma^i_{j0} = (\mathcal{H} - \dot\Phi)\delta^i_j \tag{86}$$

$$\Gamma^i_{jk} = -(\delta^i_j\partial_k + \delta^i_k\partial_j)\Phi + \delta_{jk}\partial^i\Phi, \tag{87}$$

with$\mathcal{H} = \frac{\dot{a}}{a}$.

2.6 Radiation anisotropies

We can rewrite the LHS of Eq. (53) as

$$\frac{dP^\mu}{d\lambda} = \frac{d\tau}{d\lambda}\frac{dP^\mu}{d\tau} = P^0\frac{dP^\mu}{d\tau}, \tag{88}$$

where the last equality is derived from the definition of P^μ. We can now use our perturbed Christoffel symbols to obtain the time component of the Eq. (53)

$$\frac{dP^0}{d\tau} = -(\mathcal{H}+\dot{\Psi})P^0 - 2\partial_i\Psi P^i - \left[\mathcal{H} - \dot{\Phi} - 2\mathcal{H}(\Psi+\Phi)\right]\delta_{ij}\frac{P^iP^j}{P_0}. \tag{89}$$

Question 2.11 Derive the equation above.

We aim to determine the evolution of the radiation energy density as measured by an observer. The 4-momentum here is the 4-momentum in a coordinate frame comoving with the photon in a curved spacetime. We thus have to relate this momentum to the momentum as observed in the local inertial frame of the observer (which is asymptotically flat). In an unperturbed universe these are conformally related (as we saw in Eq. (55)), in a perturbed universe we have to be more careful. Since the magnitude of the 4-momentum is frame independent, we can relate the 4-momentum in one inertial frame (associated with the metric $g_{\mu\nu}$) to another frame with metric in our local inertial frame $\eta_{\hat{\mu}\hat{\nu}}$ (note that we use ˆ for the indices in this frame to distinguish them from the coordinte frame indices):

$$\eta_{\hat{\mu}\hat{\nu}}P^{\hat{\mu}}P^{\hat{\nu}} = g_{\mu\nu}P^\mu P^\nu. \tag{90}$$

We denote the energy and the 3-momentum in the observer frame as before $P^{\hat{\nu}} = (E, P^{\hat{i}}) = (E, E\hat{p}^{\hat{i}})$. The observer is at rest and the spatial basis vectors align with the spatial coordinate directions and we can write:

$$-E^2 + \delta_{ij}P^{\hat{i}}P^{\hat{j}} = g_{00}(P^0)^2 + g_{ij}P^iP^j. \tag{91}$$

We find:

$$E = \sqrt{-g_{00}}P^0 \tag{92}$$

$$p^2 \equiv g_{ij}P^iP^j = \delta_{ij}P^{\hat{i}}P^{\hat{j}}. \tag{93}$$

Using these relations we can write:

$$P^0 = \frac{E}{\sqrt{-g_{00}}} = \frac{E}{\sqrt{a^2(1+\Psi)}} \simeq \frac{E}{a}(1-\Psi), \tag{94}$$

$$P^i = \frac{E}{\sqrt{g_{ii}}}\hat{p}^i = \frac{E}{\sqrt{a^2(1-2\Phi)}} \simeq \frac{E}{a}(1+\Phi)\hat{p}^i. \tag{95}$$

Question 2.12 Show that at linear order in perturbations

$$\frac{1}{E}\frac{dE}{d\tau} = \mathcal{H} + \dot{\Phi} - \hat{p}^i\partial_i\Psi. \tag{96}$$

Note that this indeed reduces to the unperturbed version of Eq. (57).

The first term on the RHS we aleady encountered before and captures the fact that the photons in an expanding universe loose energy, where the energy loss is proportional to the inverse scale factor (which related to redshift as $a = 1/(1+z)$). The second term on the RHS can be understood by associating Φ with a perturbation in the scale factor, i.e. $\tilde{a}(\tau, \mathbf{x}) = a(\tau)(1-\Phi)$. The last term on the RHS is due to the photon moving either into (blueshift) or out (redshift) of a gravitational potential. With some manipulation we can rewrite the equation above as

$$\frac{d\ln(aE)}{d\tau} = \frac{d\Psi}{d\tau} + \Phi' + \Psi'. \tag{97}$$

The equation above determined how perturbations in the metric affect the evolution of the radiation energy density.

To obtain an expression for the observed temperature fluctuations in the CMB as observed by an observer in a local inertial frame we integrate the equation above. First, we identify the energy with the temperature as

$$aE \propto a\bar{T}(\tau)(1+\Theta(\tau,\mathbf{x})) \tag{98}$$

and expanding the logarithm to first order, i.e.

$$\ln aE = \Theta + \mathcal{O}(\Theta^2). \tag{99}$$

Now we need to define comoving distances in our universe. The conformal time τ is related to cosmological time as

$$\tau = \int dt/a = \int \frac{da}{a^2H}. \tag{100}$$

The comoving distance is then simply $\chi = c\tau$, which if we set $c = 1$, sets $\chi = \tau$. If observations are made at $(\tau_0, \mathbf{x}_0)$, with $\mathbf{x}_0 \equiv 0$, and the photons last scattered at $(\tau_*, \mathbf{x}_*)$, with $\mathbf{x}_*$ the comoving distance , i.e. $\chi_*\hat{\mathbf{n}} = (\tau_0 - \tau_*)\hat{\mathbf{n}}$ between us and the last-scattering surface in direction $\hat{\mathbf{n}}$ on the sky [11] We can now perform the integral and obtain

$$\Theta_0(\hat{\mathbf{n}}) = \Theta_* - (\Psi_0 - \Psi_*) + \int_{\tau_*}^{\tau_0} d\tau(\dot{\Phi} + \dot{\Psi}). \tag{101}$$

We can forget about the term Ψ_0 since it presents the potential at our location which is the monopole and unobservable.

The first term on the RHS is the temperature fluctuation at last scattering. Before we move on, let us take a step back. The gravitational induced anisotropies are complemented by scattering induced anisotropies. Solving these will require the full Boltzmann equations which is beyond the scope of these lectures (you can read up on in any text book on cosmology – e.g Dodelson and Schmidt). The dominant scattering mechanism to affect CMB anisotropies is classical Thomson scattering off free electrons (with number density n_e). The evolution of the temperature fluctuation due to this scattering is given by (no derivation):

$$\frac{d\Theta}{d\tau} = -an_e\sigma_T\Theta + \frac{3an_e\sigma_T}{16\pi}\int d\hat{\mathbf{m}}(1 + (\hat{\mathbf{n}}\cdot\hat{\mathbf{m}})^2)\Theta(\hat{\mathbf{m}}) - an_e\sigma_T\hat{\mathbf{n}}\cdot\mathbf{v}_\mathrm{b} \tag{102}$$

Here σ_T is the Thomson cross-section. The first term on the RHS describes scattering out of the beam. The second term describes scattering into the beam while the final term is a Doppler effect that arises from the peculiar (bulk) velocity of the electrons, $\mathbf{v}_b = d\mathbf{x}/d\tau$. Scattering from a moving electron boosts the energy of a photon scattered in the direction that the electron is moving (see Fig 7.6 in Baumann). The scattering should be combined with the gravitational contribution. i.e. they are balanced. After some algebra it can be shown that we can write:

$$\frac{de^{-\tilde{\tau}}(\Theta + \Psi)}{d\tau} = -\dot{\tilde{\tau}}e^{\tilde{\tau}}\left(\Psi - \hat{\mathbf{n}}\cdot\mathbf{v}_b + \frac{3}{16\pi}\int d\hat{\mathbf{m}}(1 + (\hat{\mathbf{n}}\cdot\hat{\mathbf{m}})^2)\Theta\right) + e^{-\tilde{\tau}}(\dot{\Phi} + \dot{\Psi}) \tag{103}$$

Here we have defined the optical depth $\tilde{\tau} = \int an_e\sigma_T d\tau$ (sorry for the notation, we wont be discussing the optical depth beyond here). $-\dot{\tilde{\tau}}e^{\tilde{\tau}}$ is the visability function and denotes the probability that a photon last scattered at some time τ.

A solution to the above equation can be obtained by integrating along the line of sight from some early time $\tau = 0$ (where the optical depth goes to infinity) to today $\tau = \tau_0$ (when the optical depth goes to zero). The formal solution that is generated by standard CMB Boltzmann codes, such as CAMB and CLASS, via:

$$[\Theta + \Psi]_0 = \int_{\tau=0}^{\tau_0} S d\tau, \tag{104}$$

where S is the source function which is on the RHS of Eq. (103). This is an integral we can not perform analytically. Instead we can make a few simplifications.

If we neglect the anisotropic nature of Thomson scattering, we can replace $1 + (\hat{\mathbf{n}}\cdot\hat{\mathbf{m}})^2 = 1 + \cos\phi^2$, with ϕ the scattering angle, by its angular average 4/3. We also define the monopole of the temperature fluctuation as $\int \frac{d\hat{\mathbf{m}}}{4\pi}\Theta = \bar{\Theta}$. In addition, we assume that the visability function is a delta function. Then integrating yields the following expression:

$$[\Theta + \Psi]_0 = \bar{\Theta}_* + \Psi_* - \hat{\mathbf{n}}\cdot\mathbf{v}_\mathrm{b}|_* + \int_{\tau_*}^{\tau_0} d\tau(\dot{\Phi} + \dot{\Psi}) \tag{105}$$

Now again noting that Ψ_0 can not be observed, and identifying $\bar{\Theta}_*$ with the monopole at last scattering, which, given the relation between the energy density and the temperature, equates to $\delta_\gamma/4$, with δ_γ fluctuations in the radiation energy density.

Re-ordering terms, we write

$$\Theta_0(\hat{\mathbf{n}}) = [\frac{1}{4}\delta_\gamma + \Psi]_* - \hat{\mathbf{n}}\cdot\mathbf{v}_\mathrm{b}|_* + \int_{\tau_*}^{\tau_0} d\tau(\dot{\Phi} + \dot{\Psi}). \tag{106}$$

Looking back at Eq. (101) we can then equate Θ_* (the temperature fluctuations at last scattering) to those sourced by both the scattering and gravitational contributions at last scattering (here we derived them from the full equations, but in e.g. Baumann Sec 7.2.3 these terms are heuristically introduced – hence here we end up explicitly with the monopole $\bar{\Theta}$ at last scattering as opposed to Θ).

The first term on the RHS is the so called Sachs-Wolfe term. It is a combination of the intrinsic temperature fluctuations at last scattering and those induced by the gravitational redshift of the photons.

The second term on the RHS is the Doppler term. Since this term is proportional to the velocity of the fluid, while the first terms is a combination of a potential and the density (of the plasma), these terms are typically out of phase. The addition of these terms will lead to a reduction of the contrast in the troughs and peaks of the (CMB) temperature power spectrum.

The last term on the RHS is the so-called integrated Sachs-Wolfe effect. It describes an additional gravitational redhsift due to the evolution of the potentials. During matter domination in the universe, the potentials are constant and this term is zero. Since Dark Energy has become the dominant source of energy in our universe, the potentials have started to evolve, which has caused the strongest contribution of this term on large angular scales in the cosmic microwave background. There are ISW contributions from very early times which contribute mostly on small scales.

Question 2.13 Explain why late-time ISW should be expected mostly on large scales, while early-time ISW is expected to contribute more on small scales.

[11] Note that here we approximate the moment of last scattering to be instant. While this is a good approximation, in reality the surface of last scattering has a finite thickness.

2.7 CMB spectra

With an expression for the temperature fluctuations, we are now ready to compute CMB correlations functions or more importantly their spherical harmonic equivalent, the CMB spectra. We will start with the power spectrum. We will derive an expression for the CMB bispectrum. What will turn out to be the key ingredient, as eluded to earlier, are the radiation transfer functions, which linearly relate the 'primordial' fluctuations to the late-time CMB spectra. For any correlation functions of order n, we expect to appear a product of n transfer functions.

Let us start by writing the temperature fluctuations as follows:

$$\Theta(\hat{\mathbf{n}}) = \int \frac{d^3k}{(2\pi)^3} e^{i\mathbf{k}\cdot(\chi_*\hat{\mathbf{n}})} \left[A(\tau_*, \mathbf{k}) - i(\hat{\mathbf{k}}\cdot\hat{\mathbf{n}})B(\tau_*, \mathbf{k})\right], \tag{107}$$

where we have defined

$$A(\tau, \mathbf{k}) \equiv \frac{1}{4}\delta_\gamma + \Psi + \int_\tau^{\tau_0} d\tau'(\dot{\Phi} + \dot{\Psi}) \tag{108}$$

$$B(\tau, \mathbf{k}) \equiv v_b \tag{109}$$

and have used $\mathbf{v}_\mathrm{b} = i\hat{\mathbf{k}}v_b$ in Fourier space. We can factor out the initial *curvature* perturbations $\mathcal{R}_i(\mathbf{k}) = \mathcal{R}(0, \mathbf{k})$ (the curvature perturbations are related to the perturbations in the inflaton as we will see in the next lecture, we have a label $_i$ here which we will drop later on). The main advantage is that the new functions $\tilde{A}(k) \equiv A(\tau_*, \mathbf{k})/\mathcal{R}_i(\mathbf{k})$ and $\tilde{B}(k) \equiv B(\tau_*, \mathbf{k})/\mathcal{R}_i(\mathbf{k})$ depend on the magnitude k of the wavevector only because any directional dependence hidden in the perturbations (e.g. δ_γ and Ψ), must be sourced by the initial conditions.

Question 2.14 Show that

$$e^{i\mathbf{k}\cdot\chi_*\hat{\mathbf{n}}} = \sum_\ell i^\ell(2\ell+1)j_\ell(k\chi_*)P_\ell(\hat{\mathbf{k}}\cdot\hat{\mathbf{n}}), \tag{110}$$

allows us to write

$$\Theta(\hat{\mathbf{n}}) = \sum_\ell i^\ell(2\ell+1)\int \frac{d^3k}{(2\pi)^3}\Theta_\ell(k)\mathcal{R}_i(\mathbf{k})P_\ell(\hat{\mathbf{k}}\cdot\hat{\mathbf{n}}), \tag{111}$$

where we have defined:

$$\Theta_\ell(k) \equiv \tilde{A}(k)j_\ell(k\chi_*) - \tilde{B}(k)\dot{j}_\ell(k\chi_*). \tag{112}$$

Now we will use the equivalent of Eq. (18) for curvature perturbations $\mathcal{R}$[12]:

$$\langle\mathcal{R}_i(\mathbf{k})\mathcal{R}_i(\mathbf{k}')\rangle \equiv \frac{2\pi^2}{k^3}\Delta^2_\mathcal{R}(2\pi)^3\delta(\mathbf{k}+\mathbf{k}') \tag{113}$$

Using

$$P_\ell(-\hat{\mathbf{k}}\cdot\hat{\mathbf{n}}) = (-1)^\ell P_\ell(\hat{\mathbf{k}}\cdot\hat{\mathbf{n}}) \tag{114}$$

and

$$\int d\hat{\mathbf{k}}P_\ell(\hat{\mathbf{k}}\cdot\hat{\mathbf{n}})P_{\ell'}(\hat{\mathbf{k}}\cdot\hat{\mathbf{n}}') = \frac{2\pi}{2\ell+1}P_\ell(\hat{\mathbf{n}}\cdot\hat{\mathbf{n}}')\delta_{\ell\ell'} \tag{115}$$

we can write:

$$\langle\Theta(\hat{\mathbf{n}})\Theta(\hat{\mathbf{n}}')\rangle = \sum_\ell \frac{2\ell+1}{4\pi}\left[4\pi\int d\ln k\,\Theta^2_\ell(k)\Delta^2_\mathcal{R}(k)\right]P_\ell(\hat{\mathbf{n}}\cdot\hat{\mathbf{n}}') \tag{116}$$

Now going back to Eq. (27) and realizing that $\hat{\mathbf{n}}\cdot\hat{\mathbf{n}}' = \cos\theta$ with θ the angle between $\hat{\mathbf{n}}$ and $\hat{\mathbf{n}}'$ we can identify the angular temperature power spectrum:

$$C_\ell = 4\pi\int d\ln k\,\Theta^2_\ell(k)\Delta^2_\mathcal{R}(k) \tag{117}$$

The angular power spectrum is a convolution of the curvature power spectrum (which are set by inflation as we will derive in the next Lecture) and the radiation transfer functions $\Theta_\ell(k)$. These transfer functions encode both the evolution up to decoupling as well as the projection onto the last scattering surface. The temperature CMB power spectrum is shown in Fig. 3, where the best-fit models is compared to the real data. Residuals are shown in the bottom panel. By changing the initial conditions and the transfer functions, it is possible to obtain very tight constraints on cosmological parameters.

In a Boltzmann code, such as CAMB and CLASS instead of having the semi-analytical transfer functions, what is computed are the full transfer functions given by the time integral of the source functions in Eq. (104).

Assignment 2.15 Install CAMB on your laptop (if you did not bring one, please join a student who has one). Will, can we build a simple instruction or link to one? Plot the temperature power spectrum, using the best-fit Planck parameters. Play around with the values of these parameters and see how they affect the power spectrum. [check if they can separate the ISW, the SW and the Doppler terms]

[12] In Eq. (18) there is a minus sign is the momentum conserving Kronecker delta, this is because we correlated a field and its complex conjugate in that example.

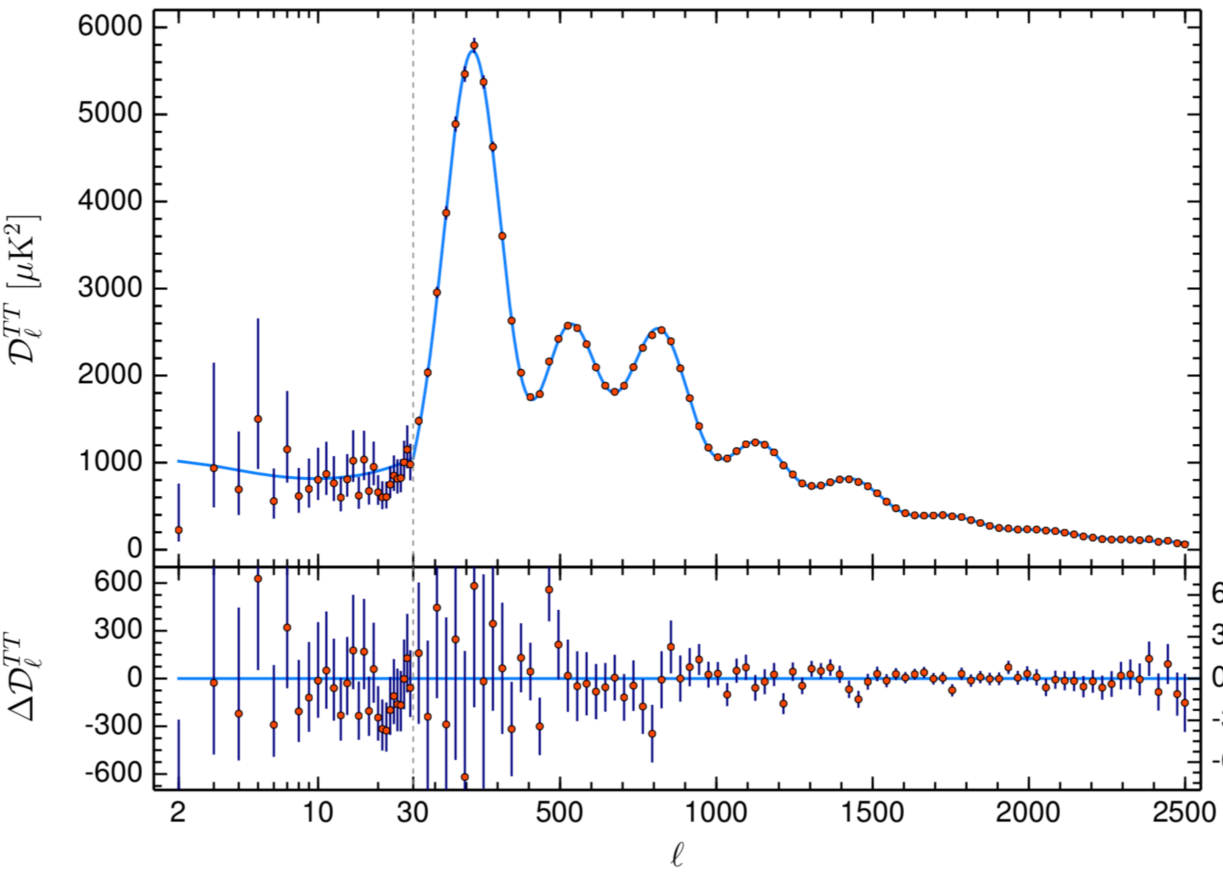

Figure 3: CMB temperature power spectrum as measured by Planck. $\mathcal{D} = \ell(\ell+1)C_\ell/2\pi$. Figure taken from `https://www.aanda.org/articles/aa/full_html/2020/09/aa33910-18/aa33910-18.html#page=1`.

Lecture 3: Inflation

In this lecture we will introduce the idea of cosmological inflation. First proposed in the early 80's by several people independently[13], it aimed to solve several apparently open conundrums in cosmology at the time, namely why the temperature of the CMB appears to be so uniform across the sky (the Horizon problem) and why the universe does not seem to have much (geometric) curvature (the flatness problem). Soon after inflation was introduced, it was shown that while the background equations describing the evolution of the field that drives inflation (the 'inflaton' field) were able solve these two conundrums (and a few others), if the inflaton is promoted to a quantum field, it can naturally produce small variations in the initial density of the universe. The statistics of these perturbations can be captured by the power spectrum (and higher order spectra if the statistics is non-Gaussian), and this power spectrum is precisely the source of observed temperature fluctuations in the CMB (as prescribed in Eq. (117)) – and eventually the source of structure in our entire universe. It should be noted that at the time inflation was postulated, the CMB had not been mapped in detail at all. For example, the absolute temperature measurements where not very accurate and measurements of the fluctuations in the CMB were mostly upper imits[14]. So while inflation was proposed to solve some problems, it also made several key prediction which were later observed and confirmed. For that reason, there is a strong consensus within the community that inflation is the most likely candidate as a source of initial conditions in our Universe. I will start by explaining the original motivations for proposing inflation. I will then provide some examples of how inflation seems to solve some aspects of our universe that were only observed long after inflation was first proposed.

3.8 Flatness Problem

The first so-called background problem is that we do not observe any spatial curvature in our universe. While I did not include curvature in our background FRLW metric, but in in Eq. (48) I wrote down how the Friedmann equation is altered. We define:

$$\Omega_k \equiv \frac{k}{H^2 a^2} \tag{118}$$

Observations tell us that Ω_k today is consistent with zero, $\Omega_{k,0} = k/H_0^2 = 0.000 \pm 0.005$[15]

[13]For a somewhat one-sided historic reconstruction, I recommend 'The Ultimate Free Lunch" by Alan Guth, one of the first people to propose inflation.

[14]for a nice video of the history of the measurements of fluctuations in the CMB see `https://www.youtube.com/watch?v=YNvL6sGLrco`

[15]There has been speculation that, depending on varying some other degrees of freedom in the standard model of cosmology, there some weak evidence for non-zero curvature, see e.g. . The flatness problem in that case would still remain.

The most general homogeneous and isotropic space can have curvature, i.e. $k \neq 0$. We can write:

$$\Omega_k' = -a'' \frac{2k}{a'^3} = \frac{6k\pi G}{3} \frac{a}{a'^3}(\rho + 3P) \tag{119}$$

where I used the second Friedmann equation Eq. (47). using the equation of state, the last line is $\propto 1 + 3\omega$. Since in an expanding universe $a' > 0$, we conclude that for a universe dominated by radiation or matter, $\Omega_k' > 0$. Consequently, for any small curvature we observe today, the positive time derivative suggests it must have been smaller in the past. This is the flatness problem.

Solving the flatness problem could be achieved by fine-tunning, for example, perhaps the curvature of the universe was exactly zero at all times (and hence it did not grow). It can be shown that is not a very appealing solution simply because we expect at least some level of fluctuations on all scales. Fluctuations of the current Hubble size would lead to a curvature of the order $\Omega_{k,0} \simeq 10^{-5}$ (see Enrico's notes, Chapter 8 for a heuristic derivation of this estimate).

Another option would be that if indeed the universe would have a larger curvature, it would become curvature dominated early on, and we would not be in the universe we observe today. This anthropic argument could be a solution, but is perhaps somewhat unsatisfying.

The last option would suggest the expansion history is modified such that Ω_k stops growing in the past. From Eq. (119) that this could happen if either a'' and $a' < 0$ or if a'' and $a' > 0$. In both cases the time derivative $\Omega_k' < 0$. The first case suggests a universe that underwent a bounce (since we know $a' > 0$ in our universe). While this has certainly been posed as one of the solutions, building a consistent model that decribes such a bounce is an open problem in cosmology. Alternatively, demanding $a'', a' > 0$ suggests that there might have been a phase of accelerated expansion in the early universe. This is the idea of cosmological inflation.

3.9 Horizon Problem

Cosmological observations of far away objects allow us to see regions in the past that are much larger than the horizon at the time. For example, the cosmic microwave background appears to have a uniform temperature accross the entire sky. When the CMB was formed the region in causal connection was much smaller then the sky we currently can observe (i.e. light has been able to reach us from regions almost 14 billion light years beyond the time the CMB was formed). Any mechanism attempting to explain the observed homogeneity in a causal way then necessarily violates causality, leading the horizon problem.

We can make the horizon problem more rigorous by within the comoving distance $\chi(a_1, a_2)$ between two generic times $a_1 = a(t_1) < a(t_2) = a_2$. Specifically, let us consider two objects on opposite directions on the sky both at a distance $z = a^{-1} - 1$. Observed from our sky, we set $a_2 = 1$. The comoving distance between these two objects would thus be $2\chi_{a,1}$. From Eq.(100) (setting $c = 1$) we find

$$\chi(a, 1) = \int_{a_1}^{1} \frac{da}{a^2 H} = \frac{1}{H_0} \frac{2}{3\omega + 1} \left[\left(1 - a^{(3\omega+1)/2} \right] \right. . \tag{120}$$

The solution is correct as long as $\omega \neq -1/3$. If we ignore the recent phase of dark energy, then $\omega > 0$ and $a \ll 1$ we find

$$2\chi(a, 1) \simeq 2 \times \frac{1}{H_0} \frac{2}{3\omega + 1} \simeq \frac{\mathcal{O}(1)}{H_0}. \tag{121}$$

The comoving distance between two distant objects on opposite directions in our sky are therefore separated by a distance $\sim H_0^{-1}$.

Next, let us compare this results with the comoving particle horizon. This is defined as the comoving distance light can have travelled since the beginning of light. For radiation, in which the expansion of the universe is decelerating, we can safely take $a \to 0$. We thus have to take the integral

$$\chi(0, a) = \int_0^a \frac{d\tilde{a}}{\tilde{a}^2 H(\tilde{a})} = \frac{1}{aH} \frac{2}{3\omega + 1} \simeq \frac{\mathcal{O}(1)}{aH} = r_H \mathcal{O}(1), \tag{122}$$

where we have defined the *comoving Hubble radius.* The physical radius as simply give by ar_H.If we assume the expansion rate universe has been decelerating since from the beginning, we find

$$\frac{\chi(a, 1)}{\chi(0, a)} \simeq 2 \frac{aH}{H_0} \simeq 2a^{-(3\omega+1)/2}. \tag{123}$$

Since we already assume $a \ll 1$, and $(3\omega + 1)/2 > 0$ (for matter and radiation), we conclude that $\frac{\chi(a,1)}{\chi(0,a)} \gg 1$. In other words, we have now shown explicitly that of we observe objects far away, we are probing scales much larger than the particle horizon at the time. This is the horizon problem.

Now, instead consider the possibilty of a phase of accelerated expansion in the past, i.e. $a'' > 0$ or $\omega < -1/3$ (which is the case for energy densities associated with e.g. dark energy and inflation). In that case we can not take the initial scale factor to $a \to 0$. Assuming this period lasts from a_i to a_f we write

$$\begin{aligned} \chi(a_i, a_f) &= \frac{1}{a_f H_f} \frac{2}{|3\omega + 1|} \left[\left(\frac{a_f}{a_i} \right)^{|3\omega+1|/2} - 1 \right] \\ &\simeq \frac{1}{a_f H_f} \frac{2}{|3\omega + 1|} \left(\frac{a_f}{a_i} \right)^{|3\omega+1|/2} . \end{aligned} \tag{124}$$

As long as $a_i \ll a_f$ we find that that $\chi(a_i, a_f) \gg r_H$, i.e. comoving Hubble radius.

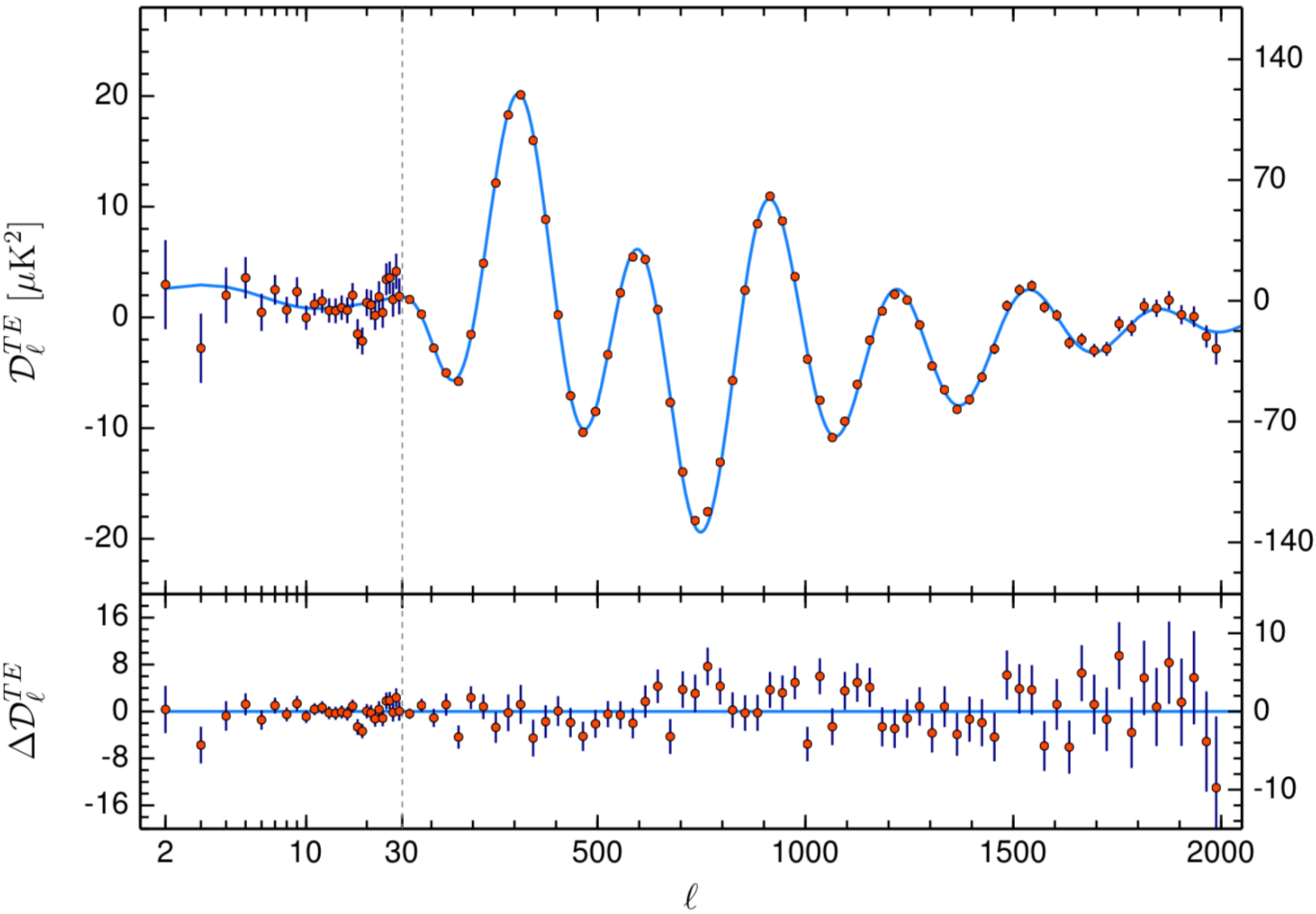

Figure 4: Temperature T and E-mode cross power spectrum. There is still coherence on very large scales. This is referred to as the coherence problem.

Question 3.16 Show that for $\omega = 1$, i.e. an equation of state for the inflaton, the particle horizon $\chi(a_i, a_f) \to 1/a_i H_i$, which is a constant.

The main conclusion here is that a phase of accelerated expansion could save the day, and would circumvent the causality issue we encountered.

The horizon and flatness problem, combined the lack of magnetic monopoles and ... where the initial motivation to propose an epoch of accelerated expansion. Only once precise measurements of the spectrum of fluctuations observed in the cosmic microwave background became available, it turned out inflation is capable of explaining several key aspects of this spectum which are hard to explain otherwise.

3.10 Phase coherence problem

By observing distance objects, we can correlate information over scales that exceeds the Hubble radius at the time. This is what we referred to as the Horizon problem. Once detailed measurements of the power spectrum of CMB fluctuations became available, it was shown that the universe also as coherent fluctuations on scales much larger then the Hubble radius at the time the CMB was formed, i.e. the CMB power spectrum shows fluctuations that oscillate in synchronicity at wavelengths $\lambda > 1/H$. How did these fluctuations become coherent? This suggests, similar to the homogeneous and isotropic background, the fluctuations were also produced during some early phase in which they were synchronised.

Let us consider the correlation between temperature and polarization E-modes (which is the curl free part of the linear polarization of the CMB photon spectrum). We can define the cross power spectrum as (similar to Eq. (26):

$$\langle a^T_{\ell m} a^E_{\ell' m'} \rangle \equiv \delta_{mm'} \delta_{\ell\ell'} C_\ell^{TE}. \tag{125}$$

The spectrum as measured by the Planck satellite is show in Fig. 4. Temperature fluctuations trace densities in the underlying photon-electron-baryon plasma. E-mode polarization traces the divergence of the plasma velocity. To build an intuitive understanding of what we should expect theoretically, we apply this simplification and write

$$\begin{aligned} T(\mathbf{x}, t) &\propto \delta(\mathbf{x}, t), \\ E(\mathbf{x}, t) &\propto \partial_i v^i(\mathbf{x}.t) \end{aligned} \tag{126}$$

Hence:

$$\langle a^T_{\ell m} a^E_{\ell m}\rangle \propto \langle \delta \partial_i v^i\rangle. \tag{127}$$

Further simplifying that we can describe the fluctuations in density fields and monochromatic sounds waves (and limiting to 1 spatial dimension):

$$\delta(x,t) A\cos(kx)\cos(\omega + \phi) \tag{128}$$

with ω some time independent frequency, A the amplitude of the wave abd ϕ the phase. For any given density perturbation, A and ϕ are some randomly drawn variables from a yet to be determined distribution. We can use the (linearized) fluid continuity equation to derive a relation between the spatial derivative of the velocity and the density fluctuations:

$$\delta' = -\partial_i v^i, \tag{129}$$

and we can thus easily determine :

$$\partial_i v^i(x,t) = \omega A\cos(kx)\sin(\omega t + \phi). \tag{130}$$

In order to estimate the correlation function in Eq. (127) we need to make some assumption about the distributions ϕ and A are drawn. It can be shown that perturbations in the CMB for scales $\ell \sim 70$ where super horizon at the time photons last scattered. For slightly smaller modes, $70 \leq \ell \leq 150$, none of these modes have been sub-horizon for more than a Hubble time. It should therefore be a reasonable assumption that on these scales, since there can have not been any causal physics creating coherence, the distribution for ϕ is flat, i.e. incoherent phases. We can then derive the corretiona function as

$$\langle \delta \partial_i v^i\rangle \propto \langle AA\rangle \int_0^{2\pi} d\phi \cos(\omega t + \phi)\sin(\omega t + \phi) = 0, \tag{131}$$

independent of the distribution of amplitudes (we could also have argued they are incoherent on the same scales). Note that non-random variables ω and $\cos(\omega t)$ can be factored out of the average. So theoretically, without any prior mechanism to make sure phases become correlated on non-causal scales, we expect the cross-correlation between temperature and E-mode polarization fluctuations to be 0 on such scales. In reality we measure a negative non-zero correlation. This again suggests that we need some mechanism early on to synchronise the phases.

3.11 Approximate scale invariance

While the CMB temperature power spectrum contains acoustic features which represent a strong scale dependence, on large enough scales the spectrum is in fact scale dependent. On sufficiently large scales the fluctuations in the CMB are dominated by the SW terms (also ISW, but these only become important on the really small scales – we will have to wait a few Hubble times for the ISW term to become the dominant term on large scales). From Eq. (106)

$$\Theta_0(\mathbf{n}) \simeq [\frac{1}{4}\delta_\gamma + \Psi]_* \tag{132}$$

The fluctuations are superhorizon during matter domination (when the CMB formed). Assuming adiabatic initial conditions and ignoring effects from Neutrinos, it can show that that (see e.g. Baumann Chapter 6)

$$[\frac{1}{4}\delta_\gamma + \Psi]_* = \frac{1}{3}\Psi_* = \frac{1}{5}\mathcal{R}_i \tag{133}$$

Hence the transfer function on large scales can be approximated as

$$\Theta^{\mathrm{SW}}_\ell(k) = \frac{1}{5} j_\ell(k\chi_*). \tag{134}$$

The observed C_ℓ on large scales as shown in Fig. xx scales as ℓ^2 (as $\ell^2 C_\ell \sim$ constant. From Eq. (117) we find in the SW limit:

$$C^{\mathrm{SW}}_\ell = \frac{4\pi}{25}\int_0^\infty d\ln k \Delta^2_{\mathcal{R}} j^2_\ell(k\chi_*). \tag{135}$$

Since $\Delta^2_{\mathcal{R}}(k)$ can only depend on the modulus of $\mathbf{k}$ (due to statistical isotropy and homogeneity as derived in the first lecture) we can make an assumption that $\Delta^2_{\mathcal{R}}(k) \propto k^{n_s - 1}$, i.e. is some power law. In that case, it can be shown that there is analytic solution for C^{S}_ℓ.

Question 3.17 Show that the analytical solution of the following integral is given by:

$$\int_0^\infty d\ln k k^{n_s-1} j^2_\ell(ak) = \frac{\sqrt{\pi} a^{1-n_s}\Gamma\left(\frac{3}{2} - \frac{n_s}{2}\right)\Gamma\left(l + \frac{n_s}{2} - \frac{1}{2}\right)}{4\Gamma\left(2 - \frac{n_s}{2}\right)\Gamma\left(l - \frac{n_s}{2} + \frac{5}{2}\right)} \tag{136}$$

Then show that for the observed power spectrum $C_\ell \propto \ell^2$ on large scales, $n_s \to 1$, i.e. the primordial power spectrum needs to be scale invariant.

Now what is required to actually obtain a scale invariant power spectrum. Mathematically it says that

$$\langle \mathcal{R}(\mathbf{x}_1)\mathcal{R}(\mathbf{x}_2)...\mathcal{R}(\mathbf{x}_n)\rangle = \langle \mathcal{R}(\lambda\mathbf{x}_1)\mathcal{R}(\lambda\mathbf{x}_2)...\mathcal{R}(\lambda\mathbf{x}_n)\rangle, \tag{137}$$

for real λ and positive integers n. If spacetime during some period in the early universe can be approximated by de Sitter:

$$ds^2 = a^2(-d\tau^2 + \delta_{ij}dx^i dx^j) \tag{138}$$

and $a^2 = (\tau H)^2$ with H =constant, spacetime has a symmetry:

$$\tau \to \lambda\tau, \quad \mathbf{x} \to \lambda\mathbf{x} \tag{139}$$

under which this spacetime is invariant. This is the so called dilation symmetry. If other non-gravitation background quantities (e.g. energy densities) depend only weakly on time, then the symmetry above is an approximate symmetry of the full theory and the correlators would also obey this symmetry. In Fourier space $\mathcal{R}(\mathbf{k},\tau) \to \mathcal{R}(\mathbf{k}/\lambda, \lambda\tau)$ and the two-point correlator takes the form:

$$\langle \mathcal{R}(\mathbf{k})\mathcal{R}(\mathbf{k}')\rangle = (2\pi)^3\delta(\mathbf{k}+\mathbf{k}')\frac{C}{k^3} \tag{140}$$

with C a constant. Comparing this to Eq. (113) implies $\Delta^2_{\mathcal{R}}(k) =$ constant, and hence $n_s = 1$, i.e. a scale independent primordial power spectrum can be obtained if the spacetime during some early epoch can be approximated by de Sitter.

3.12 Single Field Slow-Roll inflation

In order to address the conundrums discussed in the previous subsections, we already deduced requires a phase in the (very) early universe that demanded a accelerated expansion and an almost de Sitter spacetime. Recall that for a component with an equation of state $\omega \sim -1$, the Hubble radius $1/H$ is constant. Hence the comoving Hubble radius during inflation, shrinks (as $1/a$). After inflation, the comoving Hubble radius grows. One can visualize what happens during and after inflation by means of showing some fixed mode (of a given scale) across time as shown in Fig. **??**. This mode is generated inside the Hubble radius, but is stretched beyond this radius during inflation. After inflation end, the comoving Hubble radius starts growing and this mode re-enters our particle horizon. Hence any correlation structure that was imprinted during inflation, can appear to be on scales outside the particle horizon as determined from an expansion in a radiation and matter dominated universe. Today, we observe coherence and thermal equilibrium on scales $1/(a_0H_0) = 1/H_0$. Because $H \sim$ constant during inflation, the amount of expansion during inflation can be quantified by just the ratio of the scaele factor at the beginning and the end of the inflationairy epoch. We refer to this quantity as the number of E-folds:

$$N_{\rm tot} \equiv \ln(a_e/a_i) \tag{141}$$

The increase in the comoving Hubble radius during the post inflation era, must be compensated by the decrease in Hubble radius during inflation. The increase in the comoving Hubble radius in the post inflation era depends on the maximal temperature of the thermal plasma at the beginning of the hot Big Bang phase. Let us label this temperature as the reheating temperature T_R. If we now assume the universe was radiation dominated throughout its history [NOte to self this is obviously not the case, but while in cosmological time radiation was only a short period, we are dealing with the scale factor and most of the expansion was done in this epoch (I think)]. During radiation domination, since the density of the plasma scales as a^{-4}, $H \propto a^{-2}$, we can write

$$\frac{a_0H_0}{a_RH_R} = \frac{a_0}{a_R}\left(\frac{a_R}{a_0}\right)^2 = \frac{a_R}{a_0} \sim \frac{T_0}{T_R} \sim 10^{-28}\left(\frac{10^{15}\ \text{GeV}}{T_R}\right). \tag{142}$$

Note that $T_0 \sim 3$ K, i.e. the temperature of the CMB today. For convenience, we introduced the pivot energy of 10^{15} GeV. Now if we assume there is no significant expansion in the period between the end of inflation, and the start of the hot Big Bang evolution, i.e. $(a_eH_e)^{-1} \sim (a_RH_R)^{-1}$[16], we can derive a condition to explain the current observed thermal equilibrium on all scales as

$$(a_iH_i)^{-1} > (a_oH_0)^{-1} \sim 10^{28}\left(\frac{T_R}{10^{15}\ \text{GeV}}\right)(a_eH_e)^{-1}. \tag{143}$$

Since during inflation $H \sim$constant, $H_i \sim H_e$, we find

$$N_{\rm tot} > 64 + \ln\big(T_R/10^{15}\ \text{GeV}\big). \tag{144}$$

Hence we need at least 60 E-folds to deal with the horizon problem. If the reheating temperature is below 10^{15} GeV we would need fewer.

Coming back to the requirements to solve the flatness and horizon problem, we need a comving Hubble radius $1/aH$ that is shrinking with time:

$$\frac{d}{dt}(aH)^{-1} = \frac{a'H + aH'}{(aH)^2} = \frac{1}{a}(1-\epsilon). \tag{145}$$

Here ϵ is the first slow-roll parameter:

$$\epsilon \equiv \frac{H'}{H^2} = \frac{d\ln H}{dN}. \tag{146}$$

For the comoving Hubble radius to decrease we need $\epsilon < 1$. The RHS shows that the fractional change of the Hubble parameter $\Delta \ln H = \Delta H/H$ per e-folding expansion ΔN is small.

Question 3.18 Show first that by multiplying Eq. (145) with $-a'H$, the above equation translates to $a''/a > 0$ for $\epsilon > 0$, i.e. accelerated expansion.
Next use Eq. (46) and Eq. (47) to show that $\epsilon = \frac{3}{2}(1+\omega)$. In other words, $\epsilon < 1$ requires $\omega < -1/3$.

From the requirement of scale invariance, and the possible solution via a de Sitter phase in the early universe, we

[16] There can be corrections due to the physics of reheating, see e.g. [**<empty citation>**]

concluded in fact that $H \sim$ constant. In that case, $\epsilon \to 0$. In that limit we find that

$$\frac{1}{H}\frac{da^{-1}}{dt} = -\frac{1}{a} \to a = e^{Ht}, \tag{147}$$

i.e. the scale factor is growing exponentially.

> **Question 3.19** Derive the above result.

Because inflation has to end, and we no longer live in a de Sitter spacetime (although we might enter such a phase in the future when the universe becomes dark energy dominated again) ϵ is not zero, but the dilaton symmetry is slightly broken. This is the reason why inflation is ofter referred to as quasi-de-Sitter period.

In order for inflation to last long enough (sufficient e-foldings) we need ϵ to stay small enough for long enough time (otherwise the quasi-de-Sitter phase will end). We can define another slow-roll parameter based on this constraint:

$$\eta \equiv \frac{d\ln\epsilon}{dN} = \frac{\epsilon'}{H\epsilon}. \tag{148}$$

So the second slow-roll requirement is that $|\eta| < 1$. In principle we can define a tower of slow-roll parameters as:

$$\xi_{n\geq 3} \equiv \frac{d\ln\xi_{n-1}}{dN} \tag{149}$$

with $\xi_2 = \eta$. In practice, we typically have to worry only about $\xi_{n<4}$.

Since inflation requires $\omega < -1/3$, it is evident we can not use radiation or matter (which was already clear from the horizon problem). We know dark energy has $\omega = -1$ hence something similar to dark energy could suffice. In the early 80's it was shown that a time-dependent scalar field can produce an inflationairy period. Let us denote this scalar field as $\phi(t, \mathbf{x})$ [Sometimes I have t before $\mathbf{x}$ whhich is Enrico notation, while Baumann and Dodelson use the opposite–check carefully]. This field will have a kentic energy and there can be a potential. Throughout, we will assume that this scalar field dominated the energy density in the universe. For a general spacetime, the action associated with this field can be written as

$$S = \int d^4x\sqrt{-g}\left[\frac{1}{2}g^{\mu\nu}\partial_\mu\partial_\nu\phi - V(\phi)\right], \tag{150}$$

with $g \equiv \det g_{\mu\nu}$, which for a flat FRLW metric is equivalent to a^3. To obtain the background equations of motion, we need to vary the action, and set this to zero, as usual. The two dynamical degrees of freedom are ϕ and $g_{\mu\nu}$, which we assume only to depend on t (i.e. these are the unperturbed dofs – we will to perturbed quantities in the next section).

> **Question 3.20** Assuming a flat FRLW metric, derive the equation of motion for the scalar field ϕ and show that it is given by:
>
> $$\phi'' + 3H\phi' + \frac{dV}{d\phi} = 0. \tag{151}$$

By varying the inverse of the metric $g^{\mu\nu} \to g^{\mu\nu} + \delta g^{\mu\nu}$, we derive and expression for the energy-momentum tensor:

$$\delta S = -\frac{1}{2}\int d^4x\sqrt{-g}T_{\mu\nu}\delta g^{\mu\nu}, \tag{152}$$

which implies that

$$T_{\mu\nu} = \partial_\mu\phi\partial_\nu\phi - g_{\mu\nu}(\frac{1}{2}g^{\alpha\beta}\partial_\alpha\phi\partial_\beta\phi + V(\phi)), \tag{153}$$

where we used $\delta\sqrt{-g} = -\frac{1}{2}(g_{\mu\nu}\delta g^{\mu\nu})$.

> **Question 3.21** Derive the energy momentum tensor and show that for a flat FRLW metric
>
> $$\rho_\phi = \frac{1}{2}\phi'^2 + V(\phi), \tag{154}$$
>
> $$P_\phi = \frac{1}{2}\phi'^2 - V(\phi). \tag{155}$$
>
> Next use $\nabla_\mu T^{\mu\nu}$ re-derive the Klein-Gordon equation for the inflaton field Eq. (151).

The above equations for the pressure and density of the scalar field suggest that if the kinetic energy of the field ϕ'^2 is much smaller than the potential energy of the field, $\rho_\phi \simeq -P_\phi$ and $\omega \to -1$, which is the e.o.s. for a cosmological constant or dark energy. Such a configuration would thus drive period of accelerated expansion as we concluded before. So a scalar field can lead to a period of inflation, as long as the kinetic energy of the field is sufficiently small compared to its potential energy.

Using equation (Eq. (47) we can derive an expression for the time derivative of the Hubble parameter:

$$H' = -\frac{1}{2}\frac{\phi'}{M_{\rm Pl}^2}, \tag{156}$$

where we have introduced the reduced Planck mass $M_{Pl}^2 = 1/8\pi G$. We can then rewrite the first slow-roll parameter in terms of the potential as

$$\epsilon = -\frac{H'}{H^2} = \frac{\frac{3}{2}\phi'^2}{\frac{1}{2}\phi'^2 + V}, \tag{157}$$

hence for $\epsilon \ll 1$ we need the kinetic energy to be small compared to the potential. This is the reason why we refer to this type of inflation as slow-roll inflation.

For inflation to last a sufficiently long time, as before we can derive

$$\eta = 2(\epsilon - \delta), \tag{158}$$

where we defined the dimensionless acceleration per Hubble time

$$\delta \equiv -\frac{\phi''}{H\phi'} \tag{159}$$

If δ is small the second term on the lHS in Eq. (151) dominates (this term is referred to the Hubble friction term, because it acts like friction as the field is rolling down a potential). The inflation speed is determined the slope of the potential.

Question 3.22 Derive Eq. (158). The relation is such that as long as $\{\epsilon, |\delta|\} \ll 1$ we have $\{\epsilon, |\eta|\} \ll 1$, i.e. we meet the requirements for inflation to happen and for a sufficiently long time.

Next, we can apply approximations to simplify our equations of motion. Since $\epsilon \ll 1$, we know that $\phi'^2 \ll V$, and

$$H^2 \simeq \frac{V}{3M_{\rm Pl}^2}. \tag{160}$$

In addition, since $\delta \ll 1$ the e.o.m. for the finlaton simplifies to

$$3H\phi' \simeq -\frac{dV}{d\phi}. \tag{161}$$

We can then approximate

$$\epsilon \simeq \frac{M_{\rm Pl}}{2}\left(\frac{V_{,\phi}}{V}\right)^2, \tag{162}$$

where we have defined the potential slow-roll parameter and $V_{,\phi} = \frac{dV}{d\phi}$. Similarly, we can approximate

$$\delta + \epsilon \simeq M_{\rm Pl}^2 \frac{V_{,\phi\phi}}{V}, \tag{163}$$

which allows us to define the potential slow-roll parameters

$$\epsilon_V \equiv \frac{M_{\rm Pl}}{2}\left(\frac{V_{,\phi}}{V}\right)^2 \qquad \eta_V \equiv M_{\rm Pl}^2 \frac{V_{,\phi\phi}}{V}. \tag{164}$$

The relation between the Hubble and potential slow-roll parameters in the slow-roll regime is given by

$$\epsilon_V \simeq \epsilon \tag{165}$$

$$\eta_V \simeq 2\epsilon - \frac{1}{2}\eta \tag{166}$$

Question 3.23 There is an exact solution relating the potential and Hubble slow-roll parameters. These can be obtained by differentiating the Friedmann equation

$$V = (3-\epsilon)H^2 M_{\rm Pl}^2 \tag{167}$$

with respect to time and using $V' = V_{,\phi}\phi'$. Show that

$$\epsilon_V = \frac{\epsilon(\eta - 2\epsilon + 6)}{4(\epsilon-3)^2} \tag{168}$$

$$\eta_V = \frac{\eta(\eta + 2\xi_3 + 6) - 2\epsilon(5\eta + 12) + 8\epsilon^2}{4(\epsilon - 3)}. \tag{169}$$

Show that these indeed reduce to the relations in Eqs. (165) and (166) in the slow-roll regime, i.e. when we neglect higher order in slow-roll.

The total number of e-foldings can be approximated in a slow-roll regime as

$$N_{\rm tot} = \int_{a_i}^{a_e} d\ln a = \int_{t_i}^{t_e} h(t)dt = \int_{\phi_i}^{\phi_e} \frac{H}{\phi'}d\phi \simeq \int_{\phi_i}^{\phi_e} \frac{1}{\sqrt{2\epsilon_V}}\frac{|d\phi|}{M_{\rm Pl}}. \tag{170}$$

Here ϕ_i and ϕ_e are the boundary values of the field for which $\epsilon_V < 1$ (as longs as slow-roll persists). Given the derived requirement that $N_{\rm tot} \geq 60$ the above relation sets an important constraint on any (successful) inflationary model.

Assignment 3.24 Work out the slow-roll parameters for a potential of the form ϕ^n potential and show how this sets bounds on the initial field value ϕ_i given the lower number of e-folds of $N_{\rm tot} > 60$ (I have to check)

Lecture 4: Gaussianities

Now that we have shown how a scalar field can generate a period of accelerated expansion in the early universe if it is dominating the energy budget, we will demonstrate how that same model is able to produce a spectrum of initial fluctuations $\mathcal{R}_i$ (which we will label $\mathcal{R}$ for now), which match the observed CMB power spectrum on large angular scales. We will use perturbation theory and show how by lifting the scalar field to a quantum field, such a spectrum is generated quite naturally, with precisely the correct scaling behaviour. Note that in this lecture we freqenctly set $M_{\rm Pl}^2 = 1$ (untill we put it back in...).

4.13 Scalar power spectrum

To obtain the power spectrum of fluctuations sourced by inflation, demands us to consider fluctuations in the inflaton field $\phi(\mathbf{x}, t) = \bar{\phi}(t) + \delta\phi(\mathbf{x}, t)$, where we have introduced $\bar{\phi}$ to be able to distinguish this from the spatially fluctuating part of the inflaton. $\phi(t)$ presents the background field that drives the accelerated expansion. The fluctuations $\delta\phi$ are the results of promoting the inflaton field to a quantum field where fluctuations appear and disappear due to the uncertainty principle. The main goal is to obtain the equation of motion of the inflaton fluctuations $\delta\phi$. Then quantizing the theory and solving the e.o.m. and computing the correlation functions will allow us to derive the power spectrum. I suggest reading Baumann, Chapter 8 for a very thorough derivation. Here we only cover a few important steps. The main point is that we need to derive the fluctuations that survive after inflation. $\delta\phi$ no longer have meaning after the end of inflation, and we should aim to work with components that survive after inflation. As it turns out, $\mathcal{R}$ is the most convenient because it relates to fluctuations in the metric and is gauge independent.

Starting with the action for the scalar field in Eq. (150) and assuming the spatially flat gauge Eq. (75). In that gauge, we have $\psi = E = 0$ and

$$\mathcal{R} = -\mathcal{H}(v + B). \tag{171}$$

v is defined through the perturbed energy momentum tensor of Eq. (71). Using the energy-momentum tensor for a scalar field Eq. (153) and expanding to linear order in $\delta\phi$ it can be shown that for $v + B = -\delta\phi/\dot{\bar{\phi}}$. We find (to leading order)

$$\mathcal{R} = \frac{\mathcal{H}}{\dot{\bar{\phi}}}\delta\phi. \tag{172}$$

OK, question to Will. I have been trying to understand what happens in the comoving gauge, where we have $v = B = 0$. I think the above relation is unaltered, since $\mathcal{R}$ is gauge independent (and the relation between $\mathcal{R}$ and $\delta\phi$ should not be affected)

Question 4.25 Derive $v + B = -\delta\phi/\dot{\bar{\phi}}$.

To obtain the perturbed e.o.m. one can perturb the action (Lagrangian) to leading order (quadratic if we are interested in the power spectrum), which should include both variations in the metric (using the spatially flat metric) as well as the inflaton. Varying the actio with respect to the relevant degrees of freedom will lead to the e.o.m. for the perturbations. We can then quantize those perturbations find the power spectrum.

However, for future purposes it is convenient to work in a different gauge, the so called comoving gauge, for which $v + B = 0$. In this cause, the main scalar fluctuation is the curvature perturbation, i.e. $h_{ij} = a^2(1 - 2\mathcal{R})\delta_{ij}$ [Ok, also here I am not quite sure, since what happens to fluctuations in g_{00}? I was not able to deduce that $A = 0$ in this gauge—]. In the comoving gauge we completely separate the perturbations in the metric from those of the inflation field, i.e. we can set $\delta\phi = 0$ when we perturb our Action/Lagrangian. This of course simplifies a lot, since we only have to keep track of perturbations in the metric. Once we find the equations of motion for $\mathcal{R}$ solve these, and compute the power specrum (or beyond), we can come back to the relation above to obtain the power spectrum of the inflaton pertrubations $\delta\phi$

To make optimal use of this gauge, we will introduce another way to write a perturbed metric, know as the ADM formalism. In this formalism, the space and time components of the metric are split into two distinct parts. The space part previously dependent on time only through the scale factor, but now it has a non-homogeneous time dependence. For any fixed time t, the spatial part of the metric describes that of a 3 dimensional hypersurface embedded in the 4-dimensional manifold of spacetime. The entirety of spacetime can then be generated by assigning one such hypersurface for every possible t. The ADM metric is given by

$$ds^2 = -N(t, \mathbf{x})dt^2 + h_{ij}(t, \mathbf{x})\left(dx^i + N^i(t, \mathbf{x})dt\right)\left(dx^j + N^j(t, \mathbf{x})dt\right). \tag{173}$$

The metric $h_{ij}(t, \mathbf{x})$ here represents the metric of the 3-dimensional hypersurfaces induced by the 4-diminesional spacetime metric, where each slice of constant t is thus defined by one such hypersurfaces. This added dynamical space metric is not the only difference with the FLRW metric, we also see that our ADM metric contains two additional new functions, the *lapse function* $N(t, \mathbf{x})$ and the *shift vector* $N^i(t, \mathbf{x})$. The shift vector represents how the spatial hypersurfaces are deformed/changed into one another with a change in time. That is, if we take any hypersurface at time t and vary the time dt, $N^i(t, \mathbf{x})$ will tell us how space has transformed/shifted into the new hypersurface $t + dt$. The lapse function on the other hand determines how big the affect is of a change of the time coordinate. Note that if we set N to 1, N^i to the 3 dimensional 0 vector, and h_{ij} to the metric of a maximally symmetric 3-space, we recover the FLRW metric as it should.

We include gravity in our Lagrangian by addition of the Ricci scalar:

$$S = \int d^4x\sqrt{-g}\left(\frac{1}{2}R + -\frac{1}{2}g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi - V(\phi)\right), \tag{174}$$

We can express the Ricci scalar in terms of the variables of the ADM formalism by considering that the geometry of the slices making up our space-time depend on the intrinsic curvature, which corresponds to the Ricci tensor $R^{(3)}_{ij}$ of the induced 3-metric h_{ij}, and the extrinsic curvature

$$K_{ij} = \frac{1}{2N}(h'_{ij} - \nabla_i N_j - \nabla_j N_i) = \frac{E_{ij}}{N}, \tag{175}$$

through the following equation

$$R = R^{(3)} + K^{ij}K_{ij} - K^2, \tag{176}$$

where $R^{(3)} = h^{ij}R^{(3)}_{ij}$, $K = h^{ij}K_{ij}$ as before and 3-space indices are lowered and raised by the induced metric tensors h_{ij} and h^{ij} respectively[17]. Inserting this into our action along with the ADM metric tensor $g_{\mu\nu}$, we obtain

$$S = \frac{1}{2}\int d^4x\sqrt{h}N\left(R^{(3)} + K^{ij}K_{ij} - K^2 + N^{-2}(\phi' - N^i\partial_i\phi) + h^{ij}\partial_i\phi\partial_j\phi - 2V(\phi)\right). \tag{177}$$

The lapse function N and shift function N^i do not have any derivatives operating on them in the action. This makes it straightforward to define their equations of motion by varying the action. Varying the action with respect to the lapse function N results in the constraint equation

$$R^{(3)} - N^2(E^{ij}E_{ij} - E^2 + (\phi' - N^i\partial_i\phi)^2) - 2V(\phi) = 0, \tag{178}$$

while varying with respect to the shift function N^i gives us

$$\nabla_i(N^{-1}(E^i_j - E\delta^i_j)) = 0. \tag{179}$$

Question 4.26 Derive the constraint equations, Eq. (178) and Eq. (179).

We can solve for N and N^i and insert the solutions back into the action, which leaves us with an expression in which our perturbations are the only dynamic variables. In turn we can then use this action to find the equations of motion for the perturbations.

Let us write the perturbed lapse and shift function as (recall that in the unperturbed metric $N = 1$ and $N^i = 0$):

$$N = 1 + N^{(1)}(t, \mathbf{x}), \qquad N^i = \partial_i\alpha(t, \mathbf{x}). \tag{180}$$

Inserting these into (178) and (179), we can derive expressions for the evolution of the perturbations in the lapse and shift function (see Maldacena, we should check, shall I make this an exercise?)

$$N^{(1)} = -\frac{\mathcal{R}'}{H}, \qquad \partial^2\alpha(t, \mathbf{x}) = \frac{\partial^2\mathcal{R}}{H} - a^2\frac{\phi'^2}{2H^2}\mathcal{R}'. \tag{181}$$

Similarly, we obtain an expression for the perturbed Ricci scalar $R^{(3)}$ using Eq. (79)

$$R^{(3)} = \frac{e^{2\mathcal{R}}}{a^2}(4\partial^2\mathcal{R} - 2(\partial\mathcal{R})^2) \tag{182}$$

In order to find the quadratic action for $\mathcal{R}$ we can use the derived perturbed expressions above in the action and expand the action to second order. For this purpose it is not necessary to compute N or N^i to second order. The reason is that the second order term in N will be multiplying the hamiltonian constraint evaluated to zeroth order which vanishes since the zeroth order solution obeys the equations of motion. There is a similar argument for N^i. We obtain:

$$S = \frac{1}{2}\int d^4x\left(ae^{-\mathcal{R}}(1 - \frac{\mathcal{R}'}{H})(4\partial^2\mathcal{R} - 2(\partial\mathcal{R})^2 - 2a^2Ve^{-2\mathcal{R}}) + a^3e^{-3\mathcal{R}}(-6(H - \mathcal{R}')^2 + \phi'^2)\right). \tag{183}$$

Using the equations of motions of the unperturbed background, integrating by parts and expanding the exponential up to second order, the action simplifies to

$$S^{(2)} = \int d^3x d\tau a\left(\epsilon\dot{\mathcal{R}}^2 - \epsilon(\partial\mathcal{R})^2\right), \tag{184}$$

where we replaced $dt = ad\tau$. Varying this action and taking the Fourier transform we can derive the equations of motion of the momentum modes of the perturbations $\mathcal{R}$

$$\ddot{\mathcal{R}}_{\mathbf{k}} + 2\mathcal{H}\dot{\mathcal{R}}_{\mathbf{k}} + k^2\mathcal{R}_{\mathbf{k}} = 0. \tag{185}$$

We can further simplify the equations if we define a normalised field

$$f_{\mathbf{k}} \equiv z\mathcal{R}_{\mathbf{k}}, \qquad z \equiv a\sqrt{2\epsilon}, \tag{186}$$

which allows us to rewrite the equations of motion as

$$\ddot{f}_{\mathbf{k}} + \left(k^2 - \frac{\ddot{z}}{z}\right)f_{\mathbf{k}} = 0. \tag{187}$$

This is sometimes referred to as the Mukhanov-Sasaki equation. To interpret the result, let us expand $\ddot{z}/z$ in the slow-roll parameter ϵ. To leading order this gives $\ddot{z}/z = 2\mathcal{H}^2(1 + \mathcal{O}(\epsilon))$ resulting in the equations of motion

$$\ddot{f}_{\mathbf{k}} + \left(k^2 - 2\mathcal{H}^2\right)f_{\mathbf{k}} = 0. \tag{188}$$

[17] The presence of the 3-curvature already suggests we want to work with the curvature perturbations given the relation between those perturbations and the intrinsic curvature as shown in Eq. (**??**).

The solutions to the equations of motion will thus strongly depend on the ratio between k and $\mathcal{H} = aH$. Early in the inflationary epoch, aH must be very small (or the Hubble horizon $1/aH$ is very large), such that all k modes of the curvature perturbation are well within the Hubble horizon. We must thus have that during early times $k \gg aH$. In this regime:

$$\ddot{f}_{\mathbf{k}} + k^2 f_{\vec{k}} = 0, \tag{189}$$

which is just a simple harmonic oscillator equation. As inflation proceeds, aH rapidly increases and soon the wavelength of the smallest k modes (where the physical wavelength is $\lambda = a/k$) start to exceed the shrinking Hubble radius, i.e. horizon crossing. This is the moment these perturbations get frozen in, because we see that as $k \ll aH$, the equations of motion become

$$\ddot{f}_{\mathbf{k}} - 2\mathcal{H}^2 f_{\mathbf{k}} = 0, \tag{190}$$

which has a non-oscillating constant mode solution (proportional to $\mathcal{H}$).

We can find a general solution as follows. First let us write:

$$\begin{aligned} \frac{\dot{z}}{z} &= \mathcal{H}\left[1 + \frac{\eta}{2}\right] \\ \frac{\ddot{z}}{z} &= \mathcal{H}^2\left[2 - \epsilon + \frac{3}{2}\eta\right] + \mathcal{O}(\epsilon^2). \end{aligned} \tag{191}$$

using Eqs.(145) and (148)).

Question 4.27 Derive the above equations. Also show that $\epsilon = 1 - \dot{\mathcal{H}}/\mathcal{H}^2$.

Next, assuming ϵ is constant (which is an assumption we made anyways when varying the action of $\mathcal{R}$), we can integrate:

$$\mathcal{H} = -\frac{1}{\tau}(1+\epsilon). \tag{192}$$

We can now rewrite the equation of motion as

$$\ddot{f}_{\mathbf{k}} + \left(k^2 - \frac{\nu^2 - 1/4}{\tau^2}\right) f_{\mathbf{k}} = 0 \tag{193}$$

with

$$\nu \equiv \frac{3}{2} + \epsilon + \frac{1}{2}\eta. \tag{194}$$

The solution to this equation is given by

$$f_{\mathbf{k}}(\tau) = \frac{\sqrt{\pi}}{2}\sqrt{-\tau} H_\nu^{(1)}(-k\tau), \tag{195}$$

$H_\nu^{(1)}$ is a Hankel function of the first kind and we have assumed the so-called Bunch Davies vacuum to be the solution if we take $-k\tau \to -\infty$.

So far all our calculations have been in classical field theory. To be able to derive the actual statistics of our perturbations for different modes k, we will have to quantize the perturbations. We can achieve this by promoting the classical $\mathcal{R}$ to an operator $\hat{\mathcal{R}}$ by decomposing it into creation and annihilation operators $\hat{a}^\dagger_{\mathbf{k}}$ and $\hat{a}_{-\mathbf{k}}$

$$\hat{\mathcal{R}}_{\mathbf{k}} = \mathcal{R}_{\mathbf{k}}\hat{a}_{\mathbf{k}} + \mathcal{R}^*_{\mathbf{k}}\hat{a}^\dagger_{-\mathbf{k}}, \tag{196}$$

where $\mathcal{R}_{\mathbf{k}}$ still represents the classical solutions and $\hat{a}_{\mathbf{k}}$ and $\hat{a}^\dagger_{-\mathbf{k}}$ follow the standard commutation relations

$$[\hat{a}^\dagger_{\mathbf{k}}, \hat{a}^\dagger_{\mathbf{k}'}] = [\hat{a}_{\mathbf{k}}, \hat{a}_{\mathbf{k}'}] = 0, \qquad [\hat{a}^\dagger_{\mathbf{k}}, \hat{a}_{\mathbf{k}'}] = (2\pi)^3\delta^3(\mathbf{k} - \mathbf{k}'). \tag{197}$$

Using $\mathcal{R} = f/z$ and obtaining a solution for $z(\tau)$:

$$z(\tau) = z_*(\tau/\tau_*)^{\frac{1}{2}-\nu} \tag{198}$$

with τ_* some reference time which we take to be $\tau_* = -k_*^{-1}$, i.e. the moment of horizon crossing of a mode with wavenumber k_*. The two-point correlation fuction of quatized fluctuation is then given by

$$\langle \hat{\mathcal{R}}_{\mathbf{k}_1}\hat{\mathcal{R}}_{\mathbf{k}_2}\rangle = \delta^3(\mathbf{k}_1 + \mathbf{k}_2)|\mathcal{R}_k|^2 \equiv (2\pi)^3 P_{\mathcal{R}}(k)\delta^3(\mathbf{k}_1 + \mathbf{k}_2). \tag{199}$$

We defined the dimensionless power spectrum:

$$\Delta^2_{\mathcal{R}}(k) = \frac{k^3}{2\pi^2}P_{\mathcal{R}}(k) = \frac{k^3}{2\pi^2}|\mathcal{R}_k|^2 = \frac{k^3}{2\pi^2}\frac{|f_{\mathbf{k}}|^2}{z(\tau)^2}. \tag{200}$$

Question 4.28 Show that

$$\Delta^2_{\mathcal{R}}(k) = \frac{k^3}{2\pi^2}\frac{1}{2\epsilon_* M^2_{\mathrm{Pl}} a^2_*}(-k_*\tau)^{2\nu-1}\frac{\pi}{4}(-\tau)|H_\nu^{(1)}(-k\tau)|^2. \tag{201}$$

Given the followig limit:

$$\lim_{k\tau\to\infty}|H_\nu^{(1)}(-k\tau)|^2 = \frac{2^{2\nu}\Gamma(\nu)^2}{\pi^2}(-k\tau)^{-2\nu} \sim \frac{2}{\pi}(-k\tau)^{-2\nu}, \tag{202}$$

show that

$$\Delta^2_{\mathcal{R}}(k) = \frac{1}{8\pi^2\epsilon_*}\frac{H_*^2}{M_{\rm Pl}^2}\left(\frac{k}{k_*}\right)^{3-2\nu}, \tag{203}$$

with $a_* = k_*/H_*$.

The above result in the main result of this part of the lecture. Generally we write the equation as follows:

$$\Delta^2_{\mathcal{R}}(k) = A_s\left(\frac{k}{k_*}\right)^{n_s-1}, \tag{204}$$

where we have absorbed all the normalization of the spectrum into an overal parameter which we call A_s and we have degfined

$$n_s - 1 \equiv 3 - 2\nu = -2\epsilon - \eta. \tag{205}$$

Since both ϵ and η are small, $n_s \simeq 1$, i.e. the power spectrum of (curvature) fluctuations is almost scale independent. Note that this did not have to be the case. What is more, the precise requirements for inflation to happen and last long enough lead to a spectrum that is scale invariant, as observed in the temperature spectrum of the CMB.

4.14 Tensor power spectrum

So far we have ignored vector and tensor perturbations. Vector perturbations quickly decay, so we will ignore them here. Tensor perturbations however source primordial gravitational waves. They are considered one of the most robust predictions of inflation, albeit the amplitude is only weakly bound from below (so perhaps they are unobservably small). We have not included tensor perturbations in the metric nor the energy density. Adding these is straightforward (see Pajer Chapter 10, and Baumann Chapter 6 an 8). We can use the comoving gauge, where we can add tensor perturbations as follows:

$$h_{ij} = a^2[(1-2\mathcal{R})\delta_{ij} + \gamma_{ij}], \qquad \partial_i\gamma_{ij} = 0, \qquad \gamma_{ii} = 0. \tag{206}$$

Solving the shift Eq. (179) and lapse Eq. (178) to first order, with $N^i = \partial_i\alpha(t,\mathbf{x}) + N_T^i(t,\mathbf{x})$ and $\partial_i N_T^i = 0$, it can be shown that $N_T^i = 0$ (see Maldacena). Applying the comoving gauge and collecting terms quadratic in γ_{ij} we find the cubic action for tensor perturbations:

$$S_T^{(2)} = \frac{1}{8}\int d^3x d\tau a^2\left[(\dot{\gamma_{ij}})^2 - (\nabla\gamma_{ij})^2\right]. \tag{207}$$

(see Baumann page 355 for a very elegant derivation of this action). We can use rotational symmetry (cosmological principle) and allign the $z-$axis of our coordinate system with the momentum of the perturbations, i.e. $\mathbf{k} = (0,0,k)$. There are two polarization modes of the tensor perturbations denoted by $f_\times$ and f_+ (standard inflation does not induce circular polarization). We define:

$$\frac{1}{\sqrt{2}}a\gamma_{ij} \equiv \begin{pmatrix} f_+ & f_\times & 0 \\ f_\times & -f_+ & 0 \\ 0 & 0 & 0\end{pmatrix}, \tag{208}$$

and we can write:

$$S_T^{(2)} = \frac{1}{2}\sum_{\lambda=+,\times}\int d^3x d\tau\left[(\dot{f_\lambda})^2 - (\nabla f_\lambda)^2 + \frac{\ddot{a}}{a}f_\lambda^2\right], \tag{209}$$

where we used $\gamma_{ij}^2 = 4(f_+^2 + f_\times^2)/a$. This is actually identical to the second order scalar mode action, except $z \to a$. Varying the action with respect to $f_\times$ and f_+ should result in the same equations of motion (modulo $z \to a$), i.e.

$$\ddot{f_{\mathbf{k},\lambda}} + \left(k^2 - \frac{\ddot{a}}{a}\right)f_{\mathbf{k},\lambda} = 0 \tag{210}$$

The effective mass (of the Klein-Gordon equation) is given by

$$\frac{\ddot{a}}{a} = \frac{\nu^2 - 1/4}{\tau^2} \tag{211}$$

with $\nu \sim \frac{3}{2} + \epsilon$. The solution to the mode functions is equivalent to Eq. (195). We find (putting back factors of $M_{\rm Pl}^2$)

$$\Delta^2_\gamma(k) = \frac{2}{\pi^2}\frac{H_*^2}{M_{\rm Pl}^2}\left(\frac{k}{k_*}\right)^{3-2\nu}, \tag{212}$$

where we have made sure the solution to the mode functions matches the Bunch Davies vacuum for $k\tau \to -\infty$ and the above is derived in the late time approximation where $k\tau \to 0$ as before.

We can read of the spectral index of the tensor power spectrum as

$$n_T \equiv 3 - 2\nu = -2\epsilon. \tag{213}$$

Note we measure the tilt also at a fixed scale, i.e. $n_T = -2\epsilon_*$. Rewriting the spectrum as a power law:

$$\Delta^2_\gamma(k) = A_T\left(\frac{k}{k_*}\right)^{n_T}, \tag{214}$$

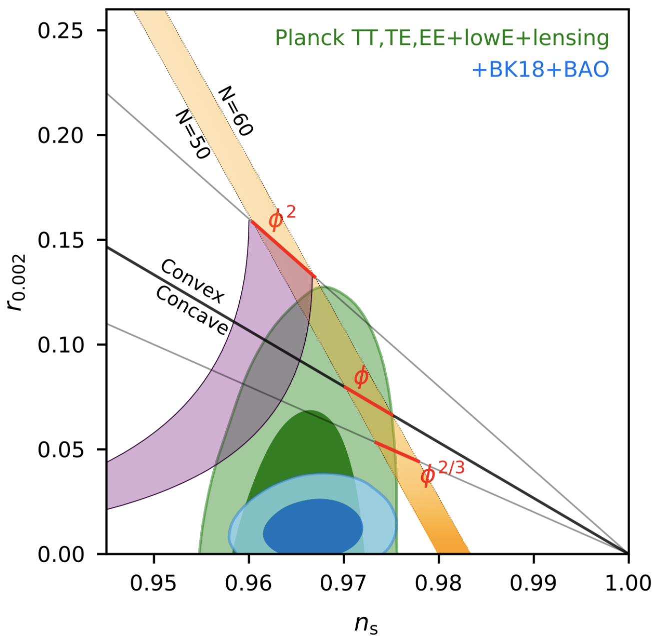

Figure 5: Figure adopted from `https://arxiv.org/pdf/2110.00483.pdf`. Contours show one and two sigma constraints on n_s and r derived from Planck, BICEP/KECK and BAO measurements. BICEP/KECK is key in putting tight constraints on r as is able to measure the B-mode pattern of the CMB with the lowest instrumental noise. Planck and BAO data are responsible for the measurement of n_s, while Planck data is used to model the dust contribution to the B modes observed with BICEP/KECK Different models of inflation are shown, with difference in the total number of e-folds. From these constraints we can conclude that power law inflation is almost ruled out. Future measurements aim to detect $r \geq 0.001$

we can associate

$$A_T \equiv \frac{2}{\pi^2}\frac{H_*^2}{M_{\rm Pl}^2}. \tag{215}$$

Question 4.29 Show that $A_T/A_s = 2\epsilon_*$. This ratio is referred to as the tensor-to-scalar ratio r and is considered one of the key targets of current and future CMB experiments. We will discuss current constraints in the next section.

The tensor-to-scalar ratio r is a metric for the amount of expansion. We can write:

$$\frac{H}{M_{\rm Pl}} = \pi\sqrt{A_s}\sqrt{\frac{r}{2}}. \tag{216}$$

i.e. the amount of expansion in units of the reduced Planck mass is proportional to the square root of r.

4.15 Observations

Now that we have derived both the tensor and scalar power spectrum we can compare these observations and discuss what these observations tell us. Since r and n_s depend on dynamics of inflation, since they are functions of the slow-roll parameters which dictate this dynamics, a key figure in observational cosmology is the n_s vs r plot.

Question 4.30 Rewrite the tensor amplitude and tilt in terms of the potential slow-roll parameters. We can define the number of e-folds remaining till the end of inflation:

$$N_* \equiv \int_{\phi_*}^{\phi_e}\frac{H}{\phi'} \simeq \int_{\phi_*}^{\phi_e}\frac{1}{\sqrt{2\epsilon_V}}\frac{|d\phi|}{M_{\rm Pl}}, \tag{217}$$

where ϕ_e is the field valid at which $\epsilon_V = 1$. Now consider a quadratic potential $m^2\phi^2$. Derive the slow-roll parameters and obtain the value of n_s and r if $N_* \sim 50$.

Recall the relation between the primordial power spectrum and the observed temperature power spectrum (explicitly labelling T for temperature and $\mathcal{R}$ for scalar):

$$C_\ell^{T,\mathcal{R}} = 4\pi\int d\ln k\left(\Theta_\ell^{T,\mathcal{R}}(k)\right)^2\Delta_\mathcal{R}^2(k). \tag{218}$$

The tensor power spectra can also affect the temperature power spectrum via:

$$C_\ell^{T,\gamma} = 4\pi\int d\ln k\left(\Theta_\ell^{T,\gamma}(k)\right)^2\Delta_\gamma^2(k), \tag{219}$$

where

$$\Theta_\ell^{T,\gamma}(k) = \frac{1}{4}\sqrt{\frac{(\ell+2)!}{(\ell-3)!}} \int_{\tau_*}^{\tau_0} d\tau \dot{\gamma}_{\pm 2}(\tau,k) \frac{j_\ell(k(\tau_0-\tau))}{k^2(\tau_0-\tau)^2}. \tag{220}$$

Here $\gamma_{\pm 2}(\tau,k) \equiv \gamma_{\pm 2}(\tau,\mathbf{k})/\gamma_{\pm 2}(0,\mathbf{k})$, i.e. the normalized mode with respect to the amplitude at the end of inflation. The tensor modes evolve according to

$$\ddot{\gamma}_{ij} + 2\mathcal{H}\dot{\gamma}_{ij} - \nabla^2\gamma_{ij} = 0, \tag{221}$$

in the absence of a source. These are related to the the two hilicities as:

$$\gamma_{ij} = \begin{pmatrix} \gamma_+ & \gamma_\times & 0 \\ \gamma_\times & -\gamma_+ & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{222}$$

and $\gamma_{\pm 2} \equiv \gamma_+ \mp i\gamma_\times$ where $+2$ is the right-handed and -2 the left-handed polarization of the gravitational wave.

This expression can be compared to Eq. (112). The main point here is that the tensor contribution tot the total observed temperature spectrum is small. Furthermore, tensors decay as the enter the horizon and the solution to Eq. (221) will show that $\gamma_{ij} \propto a^{-1}$. As a result, gravitational waves have an imprint mostly on large angular scales. On large scales, there are fewer modes to average over. This is referred to as cosmic variance. Notably, an unbiased estimator for the CMB power spectrum is given by

$$\hat{C}_\ell^X \equiv \frac{1}{2\ell+1}\sum_m |a_{\ell m}^X|^2, \tag{223}$$

with $X \in \{T, E, B\}$. We thus have $2\ell+1$ independent estimates for each m. We can determine the variance of this estimator by considering:

$$\langle a_{\ell m} a_{\ell m}^* a_{\ell m'} a_{\ell m'}^* \rangle = \langle a_{\ell m} a_{\ell m}^* \rangle \langle a_{\ell m'} a_{\ell m'}^* \rangle + \langle a_{\ell m} a_{\ell m'} \rangle \langle a_{\ell m}^* a_{\ell m'}^* \rangle + \langle a_{\ell m} a_{\ell m'}^* \rangle \langle a_{\ell m}^* a_{\ell m'} \rangle \tag{224}$$

and the signal to error ratio can be estimated as

$$\frac{C_\ell^X}{\Delta C_\ell^X} = \frac{C_\ell^X}{\sqrt{\langle (C_\ell^X - \hat{C}_\ell^X)^2 \rangle}} = \sqrt{\frac{2\ell+1}{2}}. \tag{225}$$

In other words, the unavoidable error is affecting modes disproportionally hard on low ℓ, i.e. on large scales. For the temperature power spectrum the largest contribution is not coming from gravitational waves but from the density fluctuations. To constrain gravitational waves effectively the above ratio needs to be amplified by a factor $C_\ell^{T,\gamma}/C_\ell^{T,\mathcal{R}}$, which means that the signal to error ratio is further suppressed.

The good news is that the CMB is also polarized. The polarization can be decomposed into a curl free E-mode and a divergence free B-mode. To first order, the density fluctuations only source T and E modes. At first order, gravitational waves also source B modes. Hence, the most promising observable is the CMB B-mode power spectrum on large angular scales. Several experiments are currently dedicated searching for this signal in the data. Almost all of these efforts are in discovery mode, i.e. they are targeted to observe only a small fraction of the sky for as long as possible to suppress the instrumental noise. While density fluctuations do not source B modes at linear order, their are two main observational challenges.

The first is coming from our own galaxy which is full of polarized dust. While this might lead to only a small contribution at CMB frequencies ($\sim$100 GHz), because r is so small and the B-mode amplitude sourced by gravitational waves drops quickly as function of scale, they can easily hinder a detection if not accounted for. The current adopted approach is use the frequency dependence of the dust to remove the dust from the observed data. Since the CMB frequency dependence differs to that from dust (and synchrotron at low frequncies), if there are sufficient number of channels and the spectral index of the dust does not strongly depend on the direction on the sky, it has been forecasted the dust contribution can be sufficiently supressed.

The second observational challenge is coming from the lensing of the CMB. Density fluctuations along the line of sight between us and the moment of last scattering, can deflect CMB photons in and out of the line of sight. This lensing has now been detected at many sigma by several independent experiments. What lensing also can do, is convert E modes into B modes. So while there are no primary density modes that source B modes at linear order, lensing, which is second order effect, can do this. These so-called lensing B modes are in fact the main contribution to the B modes on small angular scales. While they are suppressed on large scales, for sufficiently low noise levels, aim to detect values below current upper bounds on r, these lensing modes will obscure B modes from primordial gravitational waves. The current approach to deal with these B modes is by carefully removing using delensing; these B modes are a convolution of the density field along the line of sight and the background E modes. With knowledge of this deflection field, which can either be reconstructed from the lensed T and E (on small scales) and B modes (on small scales) or be proxied using correlated tracers, it is possible to undo the lensing, thereby removing the lensing B modes on large scales.

Currently the best constraints on n_s and r are derived from a combination of Planck data, BICEP/KECK and BAO as measured by ... The results are shown in Fig. 5. BICEP/KECK is situated on the South Pole and is build to constrain CMB polarization on large scales. With the latest data release $r_{k_*=0.05} \leq 0.035$ at 1 sigma. The best constraints on n_s are coming from Planck and BAO measurements, which tell us that $n_s = 0.9652 \pm 0.0042$ (check what pivot scale). So while we only have an upper limit on r, we have a detection of n_s, specifically one that deviates from 1 at over 3 sigma. This is by itself an important result, since the natural prediction for inflation is to have a red tilt, i.e. less power on small angular scales. The reason is the form of n_s:

$$n_s = 1 - 2\epsilon - \eta. \tag{226}$$

For slow-roll inflation $\epsilon > 0$. Now recall that $\eta = \epsilon'/H\epsilon$. In principle $\epsilon' < 0$, however this highly unlikely since for inflation to end $\epsilon \to 1$, i.e. we need $\epsilon' > 0$ to make sure this happens. All put together it is quite natural to expect $n_s < 1$ and we would have to generate a special scenario to violate this. Specifically this is signature of decreasing energy density [Ok have to make sure I am not spitting BS here].

Finally, from the observed amplitude of the CMB power spectrum we have found $A_s = 2.106 \times 10^{-9} \pm ....$ This is telling us that the typical amplitude of fluctuations is of order 10^{-5}, or in temperature units, fluctuations are of order μK on a ~ 3K background.

Let us now come back to the relation between the expansion and the tensor-to-scalar ratio. Using Eq. (??)e can write

$$\pi\sqrt{A_s}\sqrt{\frac{r}{2}} \simeq 10^{-5}\left(\frac{r}{0.01}\right)^{1/2}, \tag{227}$$

where we have used the above measurement of A_s and anticipated upper bound on r. Thus for $r \geq 0.01$ we conclude that the expansion rate during inflation was about $10^{-4}M_{\rm Pl}$. We can relate the amount of expansion to the energy scale of inflation. Recall that during inflation the energy density of a scalar field is given by

$$\rho_\phi = \frac{1}{2}\phi'^2 + V(\phi) \tag{228}$$

Therefore

$$H^2 = \frac{1}{3M_{\rm Pl}^2}\rho_\phi \simeq \frac{V}{3M_{\rm Pl}^2}, \tag{229}$$

where we used the slow-roll condition, which drives $\rho_\phi \ll V$. Since $[V] = [E^4]$ we can define:

$$E_{\rm inf} \equiv (3H^2M_{\rm Pl}^2)^{1/4} = 5 \times 10^{-3}\left(\frac{r}{0.01}\right)^{1/4} M_{\rm Pl}. \tag{230}$$

In other words, the energy scale of inflation is $\propto r^{1/4}$. With current upper bounds on r we conclude that

$$E_{\rm inf} \leq 10^{16}\ \text{GeV}. \tag{231}$$

These are really high energy scales, surpassing current terrestrial energy scales reachable in dedicated collider experiments by orders of magnitude (colliders reach ~ 10 TeV). For this reason, inflation could potentially provide insights onto physics at energy scales that are otherwise out of reach. Unfortunately, if the power spectrum of fluctuations is all we can measure, there is not that much detail we can reveal. For that purpose, we will explore statistics beyond the power spectrum in the next lecture.

Lecture 5: Non-Gaussianities

We have learned that inflation naturally produces the right initial conditions and predicts a spectrum of initial perturbations that appear to be supported by observations. The key observables of inflation are A_s, n_s, r and n_T. In principle, scale dependence of the scalar and gravitational wave power spectrum can contain more degrees of freedom, which describes the running of n_s and n_T. For n_s there is some hope that this can one day be detected even for slow-roll inflation. For the gravitational wave power spectrum, we first need to find a non-zero r. Unfortunately due to the decay of the gravitational waves, it will be really difficult to measure n_T, let alone the running of n_T. The most promising avenue there would be to detect gravitational waves from inflation using interferormeters. However, it can be shown that for a spectrum produced by inflation $n_T < 0$ (due to the weak energy condition). Hence, detecting a spectrum of primordial gravitational waves using e.g. LISA, requires gravitational waves produced through some other mechanism than standard single-field inflation (for example, by gauge fields that can source the production of classical gravitational waves which are not bound to $n_T < 0$.

Beyond these signatures, there could be isocurvature modes, which can be produces if more then one field is present during inflation. Isocurvature modes can lead to different initial perturbations in the matter and radiation waves, which will lead to distinct features in the observed spectrum of the CMB. So far there has been no evidence for isocurvature.

Several models also predict somewhat non-standard scale dependence, for example leading to sharp features or persistent oscillations in the power spectrum of initial curvature perturbations. In fact it might be quite natural to expect to deviation from a smooth power spectrum, but so far there have only been upper limits.

Instead, what would really open up a window into the early universe would be evidence for primordial non-Gaussianities. In principle, some deviations from Gaussianity are naturally expected, given that any model of the universe will have to transfer its energy eventually to know degrees of freedom, such as particles in the standard model. Since we do not expect the inflaton to couple directly to these degrees of freedom, the transfer likely goes through gravity which is a non-linear theory and will produce some non-Gaussianity. While this effect is very small for single-field slow-roll inflation, there are many other models of inflation that will predict the same power spectrum but can produce different levels and shapes of non-Gaussianity. This is exactly why primordial non-Gaussianity is still one of the most actively studied topics of inflation. Furthermore, because non-Gaussianities are proxies for non-linear interactions, if the inflaton does interact with other particles, signatures of non-Gaussianity could identify the nature of these particles, for example their masses and their spins. For that reason, the search for these non-Gaussian signals has recently been labelled as the cosmological collider experiment.

In summary, primordial non-Gaussianities are naturally produced in most models of inflation. They would open the opportunity to identify the model of inflation, perhaps providing evidence for heavy particles inaccessible through terrestrial experiment. A detection of such a signal would truly present a cosmological test of fundamental Physics.

5.16 What?

Non-Gaussianities are simply deviations from Gaussianity. As an example, consider the multivariate distribution of temperature fluctuations (Eq. (22)) in (a pixelated) Cosmic Microwave background. These can be described as

$$P(\Theta) = \frac{1}{(2\pi)^{N_{\rm pix}/2}|\xi|^{1/2}} \exp\left[-\frac{1}{2}\sum_{ij}\Theta_i(\xi)^{-1}_{ij}\Theta_j\right] \tag{232}$$

where we have denoted $\Theta_i = \Theta(\hat{\mathbf{n}})$ and $\xi_{ij} = \langle\Theta_i\Theta_j\rangle$ with $N_{\rm pix}$ the number of pixels in your CMB map. Moving to spherical harmonic space:

$$P(a) = \frac{1}{(2\pi)^{N_{\rm harm}/2}|C|^{1/2}} \exp\left[-\frac{1}{2}\sum_{\ell m\ell' m'} a^*_{\ell m}(C)^{-1}_{\ell m,\ell' m'}a_{\ell' m'}\right] \tag{233}$$

with $C_{\ell m\ell' m'} = \langle a_{\ell m}a^*_{\ell' m'}\rangle$ which as we derived equates to $\langle a^*_{\ell m}a_{\ell m}\rangle = C_\ell\delta_{\ell\ell'}\delta_{mm'}$ if we assume statistical homogeneity and isotropy. $N_{\rm harm}$ is the number of ℓ and m. In that case we have

$$P(a) = \prod_{\ell m}\frac{e^{-|a_{\ell m}|^2/(2C_\ell)}}{\sqrt{2\pi C_\ell}} \tag{234}$$

In other words, the distribution is fully specified by the covariance matrix, which in this case is C_ℓ, which is the CMB angular power spectrum. It is exactly as expected, since if we would consider a simple 1D Gaussian distribution:

$$P(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right] \tag{235}$$

and we set the mean $\mu = 0$ (which is typical in the case of perturbative quantities we are dealing with in Cosmology), we also see that this distribution is fully captured by the variance σ.

Now let us consider that is not the case. A general form for a deviation from a non-Gaussian PDF does not exist. This would require us to compute this specfically for a given source of non-Gaussianity. However, if we assume non-Gaussianity is weak, we can expand PDF (Taylor and Watts 2000):

$$\tilde{P}(a) = \mathcal{O}[P(a)] \tag{236}$$

$$\mathcal{O} = 1 - \frac{1}{3!}\sum_{\text{all } \ell m}\langle a_{\ell_1 m_1}a_{\ell_2 m_2}a_{\ell_3 m_3}\rangle\frac{\partial}{\partial a_{\ell_1 m_1}}\frac{\partial}{\partial a_{\ell_2 m_2}}\frac{\partial}{\partial a_{\ell_3 m_3}} + \text{order } a^4 \tag{237}$$

Here

$$B^{m_1 m_2 m_3}_{\ell_1\ell_2\ell_3} = \langle a_{\ell_1 m_1}a_{\ell_2 m_2}a_{\ell_3 m_3}\rangle \tag{238}$$

is the angular CMB bispectrum.

Question 5.31 Show that the modified PDF can be written as:

$$\tilde{P}(a) = P(a)\times\left[1 + \frac{1}{6}\sum_{\text{all; } \ell_i m_i} B^{m_1 m_2 m_3}_{\ell_1\ell_2\ell_3}\left(-3(C^{-1}a)_{[\ell_3 m_3}(C^{-1})_{\ell_1 m_1,\ell_2 m_2]} + (C^{-1}a)_{\ell_1 m_1}(C^{-1}a)_{\ell_2 m_2}(C^{-1}a)_{\ell_3 m_3}\right)\right] \tag{239}$$

with

$$(C^{-1}a)_{\ell m} = (C^{-1})_{\ell m,\ell' m'}a_{\ell' m'} \tag{240}$$

and

$$A_{[i_1,\ldots,i_n]} \equiv 1/n!\,[A_{i_1\ldots i_n} + \text{perm}]\,. \tag{241}$$

Because of statsitical isotropy it can be shown that

$$B^{m_1 m_2 m_3}_{\ell_1\ell_2\ell_3} = \mathcal{G}^{m_1 m_2 m_3}_{\ell_1\ell_2\ell_3}b_{\ell_1\ell_2\ell_3} \tag{242}$$

where $\mathcal{G}^{m_1 m_2 m_3}_{\ell_1\ell_2\ell_3}$ is the Gaunt factor which is just a geometrical factor from the angular integral over three spherical harmonic functions:

$$\mathcal{G}^{m_1 m_2 m_3}_{\ell_1\ell_2\ell_3} \equiv \int d^2\hat{\mathbf{n}}Y_{\ell_1 m_1}(\hat{\mathbf{n}})Y_{\ell_2 m_2}(\hat{\mathbf{n}})Y_{\ell_3 m_3}(\hat{\mathbf{n}}). \tag{243}$$

You can think of the Gaunt factor as a Dirac delta if you take the flat-sky limit for which $\ell, m \to \vec{l}$. In other words, a weakly non-Gaussian PDF can be quantified by the bispectrum.

Again this can perhaps we easier understood if we consider a 1D distribution. Suppose we modify our 1D PDF as follows:

$$\tilde{P}(x) = P^{\rm G}(x)g(x), \tag{244}$$

with $P^G(x)$ our Gaussian PDF and $g(x) \sim 1$. Suppose we aim to measure the variance of this new distribution:

$$\langle x^2\rangle = \int_{-\infty}^{\infty} x^2\tilde{P}(x)dx = \int_{-\infty}^{\infty} P^G(x)\left(1 + \left.\frac{dg}{dx}\right|_{x=0}x + \frac{1}{2!}\left.\frac{d^2g}{dx^2}\right|_{x=0}x^2 + \mathcal{O}(x^2)\right)x^2dx, \tag{245}$$

and we identify $\left.\frac{dg}{dx}\right|_{x=0} = f_{\rm NL}$ and $\left.\frac{d^2g}{dx^2}\right|_{x=0} = \mathcal{O}(f_{\rm NL}^2)$. We conclude that

$$\langle x^2 \rangle = \langle x^2 \rangle^G + \mathcal{O}(f_{\rm NL}^2). \tag{246}$$

In the Gaussian limit, i.e. when $f_{\rm NL} \sim 0$, the distribution again is specified by the variance only. In the weakly non-Gaussian limit, corrections appear that are quadratic in $f_{\rm NL}$.

> **Question 5.32** Show that $\langle x^3 \rangle = \mathcal{O}(f_{\rm NL})$, i.e. is linear in the non-Gaussian amplitude parameter $f_{\rm NL}$. In other words, it is the equivalent of the 3-point correlation function (this is called the skewness) which is most sensitive to non-Gaussianities and receives the largest correction compared to the Gaussian case.

5.17 Primordial vs Late-time

In any observed field, be that the CMB or any tracer of large scale structure, we can typically define 3 sources of non-Gaussianity:

- **Primordial non-Gaussianity**, i.e. the non-Gaussinity sources at the same time the Gaussian fluctuations were sourced.
- **Secondary non-Gaussianity.** This cover quite a lot of things in principle. Effectively it covers any source of non-Gaussianity that is changes the distribution of the true tracer field after inflation. The most evident cause of secondary non-Gaussianity is the force of gravity. In the CMB for example, assuming at last scattering the CMB photons are pristine sampled of the primordial PDF, gravitational lensing caused by matter along the line of sight between us and the last scattering surface, can deflect photons in and out of the line of sight. This effectively changes the distribution from a Gaussian one to a non-Gaussian one. For large scale structure, the most important effect is caused by gravitational collapse of the density field on small scales. This creates very large non-Gaussianities (unlike lensing, which can be treated perturbatively).
- **Foregrounds.** Any measurement of statistical properties of a tracer field, will typically be effected by stuff that is not the tracer itself, but something that is between us and the tracer. Collectively I would denote these as foregrounds. There is no reason to think these foregrounds are Gaussian. For example, for the CMB we gave large galactic foregrounds, which are particularly important for the detection of primordial B modes. These polarized foregrounds follow the morphology of the filaments and gas clouds in our own galaxy and are highly non-Gaussian.

So if we consider the CMB we could write:

$$\Theta(\hat{\mathbf{n}}) = \Theta^{\rm prim}(\hat{\mathbf{n}}) + \Theta^{\rm sec}(\hat{\mathbf{n}}) + \Theta^{\rm fgs}(\hat{\mathbf{n}}) \tag{247}$$

There are a couple of other comments to be made. First, a primordial signal can be predicted, but its amplitude and shape (as we will explain later) are a-priori unknown. In fact, it is precisely the goal to measure and interpret any primordial signal as a observational signature of fundamental physics (i.e. the title of this lecture series!). While non-Gaussianities due to foregrounds are perhaps harder to predict, the can be modelled and we can perform complementary observations (for example by considering different tracers of the same foregrounds). Secondary non-Gaussianities are sourced by known physics and can be modelled using both analytical methods and simulations[18].

Second, non-primordial non-Gaussianities are much larger than anything primordial. Because these are driven by non-linear evolution, late-time non-Gaussianities, who have had a very long time to evolve, can be large. Specifically on small scales, were interaction times are $\ll 1/H$, we see very large non-Gaussianities. However large though, this does perhaps provide some way to disentangle these effects, since in particular secondary non-Gaussianities change as a function of redshift. Instead, the primordial signal is frozen into the underlying distribution.

Finally,in this discussion we strictly separated the signal of the components, i.e. in some sense we assume that e.g.

$$\langle \Theta^{\rm prim}\Theta^{\rm sec}\Theta^{\rm sec} \rangle = 0. \tag{248}$$

5.18 Primordial sources: inflation

Even for the simplest model of inflation, i.e. single-field slow-roll, where the fields is only weakly self interacting due to gravity, we should expect some non-Gaussianity. Before we calculate this, we can get some idea of the type of signal you could expect when doing a detailed calculation. We know from the simple 1D equivalent above that for a Gaussian field the bispectrum is zero. Let us now assume that the non-Gaussian field can be expanded as follows:

$$\mathcal{R}_{\mathbf{k}} = \mathcal{R}_{\mathbf{k}}^G + f_{\rm NL}\mathcal{R}_{\mathbf{k}}^2. \tag{249}$$

Here $\mathcal{R}_{\mathbf{k}}^G$ is the Gaussian field for which the bispectrum vanishes.

> **Question 5.33** Suppose the above represents the solution of the e.o.m. for $\mathcal{R}$. By using that $\langle \mathcal{R}_{\mathbf{k}}^G \mathcal{R}_{\mathbf{k}'}^G \rangle \propto \delta(\mathbf{k} + \mathbf{k}')P_{\mathcal{R}}(k)$, show that
>
> $$\langle \mathcal{R}_{\mathbf{k}_1}\mathcal{R}_{\mathbf{k}_2}\mathcal{R}_{\mathbf{k}_3} \rangle \propto f_{\rm NL}\delta(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3)\left[P_{\mathcal{R}}(k_1)P_{\mathcal{R}}(k_2) + P_{\mathcal{R}}(k_2)P_{\mathcal{R}}(k_3) + P_{\mathcal{R}}(k_3)P_{\mathcal{R}}(k_1)\right]. \tag{250}$$
>
> This type of bispectrum is in fact referred to as the local bispectrum, which has a shape that peaks for squeezed momenta, i.e. when $k_i \ll k_j, k_m$.

[18] There are maybe some exceptions where there is more uncertainty, for example if secondary effects are caused by magnetic fields or complex baryonic feedback.

Maldacena showed in nominal paper that primordial non-Gaussianities for single-field slow-roll are slow-roll suppressed. Starting from the action in (174), and deriving the cubic action using the ADM formalism, he was able to show that $f_{\rm NL} \propto (n_s - 1)$. Several years later, Chen et al expanded that idea to include non-Canonical kinetic terms. The action for the collection of these models can be written as:

$$S = \frac{1}{2}\int d^4x\sqrt{g}\,[R + 2P(X,\phi)]\,, \tag{251}$$

where $X = -\frac{1}{2}g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi$. So for single-field slow-roll we would have $P(X,\phi) = X - V(\phi)$.

The recipe for obtaining the bispectra is then again to derive the action up to cubic order. Note again that rewriting the inflation action using the ADM metric as is designed so that one can think h_{ij} in the and ϕ as the dynamical variables and N and N^i as Lagrange multipliers. The advantage is that when you want to compute the n-point correlation function, you only need to compute the action to $n^{P\text{th}}$ order in perturbations in h_{ij} and ϕ while terms containing higher order in N and N^i will vanish.

Next you derive the interaction Hamiltonian from the perturbed action via $S = -\int d^4x H_I$. In a cosmological setting, we need to consider the time evolution of equal time correlation functions:

$$i\frac{d|\psi(\tau)\rangle}{d\tau} = H_{\rm I}|\psi(\tau)\rangle, \tag{252}$$

which has solution:

$$|\psi(\tau)\rangle = Te^{-i\int_{\tau_0}^{\tau} H_{\rm I}(\tau')d\tau'}|\psi(\tau_0)\rangle. \tag{253}$$

We can expand the exponential to arbitrary order in H_I. The correlation function for the dynamical degree of freedom (assuming the comoving gauge) $\mathcal{R}$ is then given by

$$\begin{aligned}\langle\Psi(\tau)|\mathcal{R}_{\mathbf{k}_1}(\tau)\mathcal{R}_{\mathbf{k}_2}(\tau)\mathcal{R}_{\mathbf{k}_3}(\tau)|\Psi(\tau)\rangle &= \langle\Psi(\tau_0)|\mathcal{R}_{\mathbf{k}_1}(\tau)\mathcal{R}_{\mathbf{k}_2}(\tau)\mathcal{R}_{\mathbf{k}_3}(\tau)|\Psi(\tau_0)\rangle \\ &\quad -i\int_{\tau_0}^{\tau} d\tau'\langle\Psi(\tau_0)|[\mathcal{R}_{\mathbf{k}_1}(\tau')\mathcal{R}_{\mathbf{k}_2}(\tau')\mathcal{R}_{\mathbf{k}_3}(\tau'), H_{\rm I}(\tau')]|\Psi(\tau_0)\rangle + \mathcal{O}(H_I^2).\end{aligned} \tag{254}$$

The first term on the right is zero if the initial $\Psi(\tau_0)$ is Gaussian (see for e.g. Angullo and Parker 2010 for example where the effects of mixed initial states). The second term on the RHS is the lowest order effect. Applying the reality condition we find:

$$\langle\mathcal{R}_{\mathbf{k}_1}(\tau)\mathcal{R}_{\mathbf{k}_2}(\tau)\mathcal{R}_{\mathbf{k}_3}(\tau)\rangle \simeq -2\mathcal{R}\text{e}\left[-i\int_{\tau_0}^{\tau} d\tau'\langle\mathcal{R}_{\mathbf{k}_1}(\tau')\mathcal{R}_{\mathbf{k}_2}(\tau')\mathcal{R}_{\mathbf{k}_3}(\tau')H_{\rm I}(\tau')\rangle\right]. \tag{255}$$

In other words, once we have determined H_I in terms of $\mathcal{R}$ we can compute the bispectrum of $\mathcal{R}$ by computing the integral above. Since we want to compute the spectrum at the end of inflation, and we know fluctuations are frozen once they exit the horizon (i.e. when $k = aH$) we can take $\tau \to 0$. Assuming a Bunch Davies initial state we can also set $\tau_0 \to -\infty$.

Question 5.34 Proof Eq. (255).

For the action of Eq. (251) the cubic action can be derived to be (see Chen 2010, Chen et al 2006):

$$S^{(3)} = \int d^3x d\tau a^2\epsilon\left[\frac{1}{\mathcal{H}c_s^2}\left(\left(1-\frac{1}{c_s^2}\right)+\frac{2\lambda}{\Sigma}\right)\dot{\mathcal{R}}^3 - \frac{3}{c_s^2}\left(1-\frac{1}{c_s^2}\right)\mathcal{R}\dot{\mathcal{R}}^2 + \left(1-\frac{1}{c_s^2}\right)\mathcal{R}(\partial\mathcal{R})^2 ...\right], \tag{256}$$

where ... represents terms that are subleading. We have introduced new variables:

$$c_s^2 = \frac{P_{,X}}{P_{,X} + 2XP_{,XX}}, \tag{257}$$

$$\Sigma = XP_{,X} + 2X^2P_{,XX}, \tag{258}$$

$$\lambda = X^2P_{,XX} + \frac{2}{3}X^3P_{,XXX} \tag{259}$$

As is evident these all represent (dimensionless) differential properties of the kinetic term in the action. The normalized first derivative is a proxy for the 'speed of sound', i.e. the speed of propagation of the scalar degree of freedom (one could also define one for the graviton in principle). For a canonical kinetic term $c_s = c = 1$ (in SR units). Hence in the above cubic action, terms proportional to $1 - 1/c_s^2$ are perturbative.

Question 5.35 Show that $\Sigma = H^2\epsilon/c_s^2$.

For a general action as the one above, we can derive the bispectrum of all terms. This is a little beyond the scope of this lecture series, but I encourage you to take a look at the papers mentioned above in case you are interested in those derivations. The general prediction for a bispectrum (to connect with observations, See Pimentel et al 2022) we write

$$\langle\mathcal{R}_{\mathbf{k}_1}\mathcal{R}_{\mathbf{k}_2}\mathcal{R}_{\mathbf{k}_3}\rangle = (2\pi)^3\delta^{(3)}(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3)\frac{18}{5}f_{\rm NL}^{\rm type}A_s^2\frac{S^{\rm type}(k_1,k_2,k_3)}{k_1^2k_2^2k_3^2}. \tag{260}$$

Here we have associated $\tilde{\Delta}^2_{\mathcal{R}} = \Delta^2_{\mathcal{R}}(n_s = 1)/c_s$ and $A_s = \Delta^2_{\mathcal{R}}(n_s = 1)$. The first two terms $S^{c_s,\lambda}$ are the result of the terms in Eq. (256) and are given by:

$$S^{\lambda} = \frac{3k_2k_2k_3}{2k_t^3} \tag{261}$$

$$S^{c_s} = \left(-\frac{1}{k_t}\sum_{i>j}k_i^2k_j^2 + \frac{1}{2k_t^2}\sum_{i\neq j}k_i^2k_j^2 + \frac{1}{8}\sum k_i^3\right)\frac{1}{k_1k_2k_3}, \tag{262}$$

where we have introduced $k_t = \sum_i k_i$. S^{η} and S^{ϵ} are terms that will remain after we set $c_s = 1$ and $\lambda/\Sigma = 0$ and are the usual slow-roll predictions where $f_{\rm NL} = \mathcal{O}(\epsilon)$. We can now identify an amplitude $f_{\rm NL}^{\rm type}$ associated with each shape $S^{\rm type}$. The associated amplitudes for this example are given by

$$f_{\rm NL}^{\lambda} = \frac{1}{18}\left(1 - \frac{1}{c_s^2} + \frac{2\lambda}{\sigma}\right), \tag{263}$$

$$f_{\rm NL}^{c_s} = -\frac{7}{24}\left(1 - \frac{1}{c_s^2}\right). \tag{264}$$

Large non-Gaussianities can thus arise in this (general) non-canonical model for $c_s \ll 1$ and $\lambda/\Sigma \gg 0$. While S^{λ,c_s} differ in their exact momentum dependence, the overall trend here is that the shapes peak when $k_1 \simeq k_2 \simeq k_3$, i.e. the equilateral limit. This is the opposite of the canonical single-field predictions, which has a contribution that peaks in the squeezed limit. The main point here is that the shape of the spectrum will allow us to identify of the action associated with the field driving inflation is single-field slow-roll or contains additional higher derivative interactions. More generally, a detection of the equilateral type would suggest the field is not weakly interacting. A more formal way of distinguishing different shapes is by considering the so-called soft limit, i.e. when one of the momenta $q \to 0$. Let us write:

$$F(k_1, k_2, k_3) \equiv \frac{S(k_1, k_2, k_3)}{(k_1k_2k_3)^2}. \tag{265}$$

In the soft limit:

$$F(k_1, k_2, q \to 0) = \frac{1}{k_1^3k_2^3}\left(\frac{k_1}{k_2}\right)^{\Delta}. \tag{266}$$

Let us summerize some important shapes:

- **Equilateral/orthogonal pnGs**: Bispectra that peak when $k_1 \sim k_2 \sim k_3$. The equilateral shape can be written as

$$S^{\rm equil} \propto \frac{g(k_1, k_2, k_3)}{h(k_t)} \tag{267}$$

  with h and g both functions in the same order of k.
- **Local pnGs**: Bispectra that peak when $k_1 \ll k_2, k_3$. Large local non-Gaussianities indicate presence of more than one light degree of freedom (i.e. multiple field inflation). The local shape can be written as

$$S^{\rm local} \propto k_1^2k_2^2k_3^2\left[P_{\mathcal{R}}(k_1)P_{\mathcal{R}}(k_2) + \text{perms}\right]. \tag{268}$$

- **Cosmological collider**: Due to presence of massive particles interacting with the inflaton. In the soft limit $S \sim q^{\alpha}P_s(\cos\Theta)$, with Θ the angle between soft momentum $\mathbf{q}$ and one of the hard momenta. α is either a real number $\alpha \in [-1, \frac{1}{2}]$ or given by $\alpha = \frac{1}{2} + i\mu$ where μ depends on the mass of the heavy particle. P_s is a Legendre polynomial of degree s, where s is associated with the spin of the particle. There is a clear analogy with how mass/spin are measured in particle colliders, hence the cosmological collider.

> **Question 5.36** Derive Δ, i.e the scaling in the soft limit, for each of these shapes. Do the same for the shapes in Eq. (261) and Eq. (262).

While the above methods by expanding the action to the desired order allows us to derive the bispectrum from any model (with a well defined action) there is a very elegant and quick way to derive the bispectrum for multiple-field inflation.

We can think about the curvature like regions that expand a little more then others. Another proxy of the expansion is the number of e-folds, and hence we write:

$$\mathcal{R}(t, \mathbf{x}) = \delta N = N(t, \mathbf{x}) - N(t). \tag{269}$$

Next, we write:

$$\mathcal{R} = N(\phi_a + \delta\phi_a) - N(\phi_a) \tag{270}$$

and Taylor expand:

$$\mathcal{R} = N_a\delta\phi_a + \frac{1}{2}N_{\rm ab}\delta\phi_a\delta\phi_b + ... \tag{271}$$

Here $N_a \equiv \frac{\partial N}{\partial \phi_{a_*}}$ which only depends only on background equations.

> **Question 5.37** Show that $\langle \mathcal{R}_{\mathbf{k}}\mathcal{R}_{\mathbf{k}'}\rangle = N_aN_bP_{\phi}$, with $\langle\delta\phi_a\delta\phi_b\rangle = (2\pi)^3\delta^{(3)}(\mathbf{k}+\mathbf{k}')\delta_{ab}P_{\phi}(k)$ and $P_{\phi}(k) = \left(\frac{H_*}{2\pi}\right)^2$.

Similarly, one can compute the bispectrum:

$$\langle \mathcal{R}_{\mathbf{k}_1}\mathcal{R}_{\mathbf{k}_2}\mathcal{R}_{\mathbf{k}_3}\rangle = \frac{1}{2}N_aN_bN_{cd}\langle\delta\phi_a(\mathbf{k}_1)\delta\phi_b(\mathbf{k}_2)\int\frac{d^3\mathbf{q}}{(2\pi)^3}\delta\phi_c((\mathbf{q}))\delta\phi_d(\mathbf{k}_3-\mathbf{q})\rangle + 2 \text{ perms} \tag{272}$$

After some massaging, we find

$$\langle \mathcal{R}_{\mathbf{k}_1}\mathcal{R}_{\mathbf{k}_2}\mathcal{R}_{\mathbf{k}_3}\rangle = (2\pi)^3\left[\frac{1}{2}\frac{N_aN_bN_{ab}}{(N_cN_c)^2}P_{\mathcal{R}}(k_1)P_{\mathcal{R}}(k_2) + 2 \text{ perms}\right]\delta^{(3)}(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3). \tag{273}$$

This is exactly what you should expect with $f_{\rm NL}^{\rm loc} = \frac{5}{6}\frac{N_aN_bN_{ab}}{(N_cN_c)^2}$.

5.19 Observations

Now that we have shown the primordial non-Gaussianties can reveal details about the early universe that are not distinguishable by simply measuring the power spectrum, we can start discussing how to observe this. Because of linear transfer, has so far been the most useful observable. Limiting ourselves to the bispectrum, and to the CMB for now, we specify the bispectrum:

$$\langle a_{\ell_1m_1}a_{\ell_2m_2}a_{\ell_3m_3}\rangle \equiv B^{m_1m_2m_3}_{\ell_1\ell_2\ell_3} = \mathcal{G}^{m_1m_2m_3}_{\ell_1\ell_2\ell_3}b_{\ell_1\ell_2\ell_3} \tag{274}$$

with the Gaunt factor given in Eq. (243). We can actually write down the expression for the $a_{\ell m}$ for temperature and E mode fluctuations[19]:

$$a^T_{\ell m} = 4\pi(-i)^\ell\int\frac{d^3\mathbf{k}}{(2\pi)^3}\mathcal{R}_{\mathbf{k}}\Theta^{T,\mathcal{R}}_\ell(\mathbf{k})Y^*_{\ell m}(\hat{k}) \tag{275}$$

$$a^E_{\ell m} = 4\pi(-i)^\ell\sqrt{\frac{(\ell+2)!}{(\ell-2)!}}\int\frac{d^3\mathbf{k}}{(2\pi)^3}\mathcal{R}_{\mathbf{k}}\Theta^{E,\mathcal{R}}_\ell(\mathbf{k})Y^*_{\ell m}(\hat{k}). \tag{276}$$

With these definitions we can obtain the relation between the reduced CMB bispectrum and primordial (scalar) bispectrum:

$$b^{TTT}_{\ell_1\ell_2\ell_3} = f^{\rm type}_{\rm NL}A_s^2\left(\frac{2}{\pi}\right)^3\int d\chi\chi^2\left[\prod_{i=1}^3\int dk_ij_{\ell_i}(k_i\chi)\Theta^{T,\mathcal{R}}(k_i)\right]S^{\rm type}(k_1,k_2,k_3) \tag{277}$$

and similar for bispectra that mix T and E. Here With $j_\ell(k\chi)$ spherical Bessel functions and χ conformal distance. The equation above is a 4D coupled integral. To connect to data, we have to compare each ℓ_1,ℓ_2,ℓ_3 triplet to the data, which can be computationally challenging. It was realised that if $S^{\rm type}(k_1,k_2,k_3) = \prod_i^3 f_i(k_i)$, i.e. is factorizable, this would greatly reduce computational costs. The local shape is already factorized. Templates for orthogonal and equilateral triangles (with the correct soft limit) have been proposed:

$$S^{\rm equil}(k_1,k_2,k_3) = \frac{\prod_{{\rm cyclic}\ pqr}(k_p+k_q-k_r)}{k_2k_2k_3} \tag{278}$$

$$S^{\rm ortho}(k_1,k_2,k_3) = (1+p)S^{\rm equil}(k_1,k_2,k_3) - p\frac{\Gamma(k_1,k_2,k_3)^3}{k_1k_2k_3} \tag{279}$$

with $p \simeq 8.52$ and

$$\Gamma(k_1,k_2,k_3) \equiv \frac{2}{3}\sum_{a<b}^3 k_ak_b - \frac{1}{3}\sum_a^3 k_a^2. \tag{280}$$

In principle, factorisation is not required, it is just a way too speed up calculations (i.e. if you have sufficient computing power you could do it). For shapes that are not factorized, Fergusson and Shellard (2006++) adopted a method to expand any shape using a set of carefully constructed basis functions referred to as the modal method. In Muchmeyer et al 2014 we further expanded on this method to include bispectra that contain features which is currently the only existing method that can constrain highly oscillating bispectra such as those predicted by axion-monodromy inflation. In practice, a computational speed-up can also be achieved by binning in multiple (binned bispectrum, van Tent, Bucher et al 2015). Will Coulton has a code that computes the binned bispectrum.

The original method is referred to as the KSW (Komatsu-Spergel-Wandelt) estimator. So in all, in the Planck paper we applied 3 methods: KSW, Modal, Binned which we also plan to adopt for data coming from the Simons Oservatory.

Now we have to derive an optimal estimator (i.e. one that is unbiased and has minimal variance). Recall the non-Gaussian PDF:

$$\tilde{P}(a) = P(a)\times\left[1+\frac{1}{6}\sum_{{\rm all;}\ \ell_im_i}B^{m_1m_2m_3}_{\ell_1\ell_2\ell_3}\left(-3(C^{-1}a)_{[\ell_3m_3}(C^{-1})_{\ell_1m_1,\ell_2m_2]} + (C^{-1}a)_{\ell_1m_1}(C^{-1}a)_{\ell_2m_2}(C^{-1}a)_{\ell_3m_3}\right)\right]. \tag{281}$$

We can maximize this PDF w.r.t. $f^{\rm type}_{\rm NL}$, i.e. $d\tilde{P}/df^{\rm type}_{\rm NL} = 0$. Let us write this optimal estimator as:

$$\hat{f}^{(a)}_{\rm NL} = \sum_b(F^{-1})_{ab}\mathcal{S}_b, \tag{282}$$

where ere a and b now stand for different types (e.g. local and equilateral). It is the observed and integrated equivalent

[19] When actually computing these, one might have to worry about a factor of 3/5, which sets the relation between the transfer function associated with the gravitational potential and the curvature, which differ by a factor of 3/5; here we assume these are absorbed.

of $S^{\rm type}$. Next we define $\mathcal{S}_b$:

$$\mathcal{S}_b = \frac{1}{6}\sum_{\text{all; }\ell_i m_i} \mathcal{G}^{m_1m_2m_3}_{\ell_1\ell_2\ell_3} b^{(b)}_{\ell_1\ell_2\ell_3}\left(-3(C^{-1}a)_{[\ell_3m_3}(C^{-1})_{\ell_1m_1,\ell_2m_2]} + (C^{-1}a)_{\ell_1m_1}(C^{-1}a)_{\ell_2m_2}(C^{-1}a)_{\ell_3m_3}\right) \tag{283}$$

In that case F_{ab} is the Fisher matrix for $f^{(a)}_{\rm NL}$. The covariance matrix for $f^{(a)}_{\rm NL}$ is the inverse Fisher matrix:

$$(F^{-1})_{ab} = \langle f^{(a)}_{\rm NL} f^{(b)}_{\rm NL}\rangle - \langle f^{(a)}_{\rm NL}\rangle\langle f^{(b)}_{\rm NL}\rangle \tag{284}$$

and the 1 sigma error is given by $\Delta\hat{f}^{(a)}_{\rm NL} = (F^{-1})_{aa}$. Using the definition of the Gaunt integral we can rewrite $\mathcal{S}_b$:

$$\mathcal{S}_b = \frac{1}{6}\int d^2\hat{\mathbf{n}} \sum_{\ell_1\ell_2\ell_3} b^{(b)}_{\ell_1\ell_2\ell_3}\left[e_{\ell_1}(\hat{n})e_{\ell_2}(\hat{n})e_{\ell_3}(\hat{n}) - 3d_{[\ell_1\ell_2}(\hat{n})e_{\ell_3]}(\hat{n})\right], \tag{285}$$

where

$$e_\ell(\hat{n}) = \sum_m (C^{-1}a)_{\ell m} Y_{\ell m}(\hat{n}) \tag{286}$$

$$d_{\ell\ell'}(\hat{n}) = \sum_{m,m'} (C^{-1})_{\ell m,\ell'm'} Y_{\ell m}(\hat{n}) Y_{\ell'm'}(\hat{n}) \tag{287}$$

The advantage of writing it in this form is that the summation over m can be done using an FFT (part of HEALPIX). Furthermore, because $d_{\ell\ell'}(\hat{n}) = \langle e_\ell(\hat{n})e_{\ell'}(\hat{n})\rangle$ we can compute $d_{\ell\ell'}$ by computing a (large number) of monte Carlo relizations of the CMB + noise, i.e.

$$d_{\ell\ell'}(\hat{n}) = \langle e_\ell(\hat{n})e_{\ell'}(\hat{n})\rangle_{\rm MC}. \tag{288}$$

The final expression is then:

$$\mathcal{S}_b = \frac{1}{6}\int d^2\hat{\mathbf{n}} \sum_{\ell_1\ell_2\ell_3} b^{(b)}_{\ell_1\ell_2\ell_3}\left[e_{\ell_1}(\hat{n})e_{\ell_2}(\hat{n})e_{\ell_3}(\hat{n}) - 3e_{\ell_3}(\hat{n})\langle e_{\ell_1}(\hat{n})e_{\ell_2}(\hat{n})\rangle_{\rm MC}\right] \tag{289}$$

We have dropped the symmetry brackets here, since the symmetry of the bispectrum implies we can only use this term from here on, as long as we only sum over the correct ℓ (factorials cancel the prefactor). The final term is often referred to as the linear term, since it is linear in e_ℓ. Historically this term was forgotten in the earliest analysis, artificially blowing up the error on $\hat{f}_{\rm NL}$. The integral over $\hat{\mathbf{n}}$ has to be performed over the full-sky, including masks.

The full expression of the Fisher matrix is given by:

$$F_{ab} = \frac{f_{\rm sky}}{6}\sum_{\text{all}\ell m,\ell'm'} \mathcal{G}^{m_1m_2m_3}_{\ell_1\ell_2\ell_3} b^{(a)}_{\ell_1\ell_2\ell_3}(C)^{-1}_{\ell_1m_1,\ell'_1m'_1}(C)^{-1}_{\ell_2m_2,\ell'_2m'_2}(C)^{-1}_{\ell_3m_3,\ell'_3m'_3}\mathcal{G}^{m'_1m'_2m'_3}_{\ell'_1\ell'_2\ell'_3} b^{(b)}_{\ell'_1\ell'_2\ell'_3}. \tag{290}$$

Here $f_{\rm sky}$ is the fraction of sky outside of the mask (or in case of ground based, mask + unobserved parts of sky). If the covariance matrices are diagonal we can sum mode by mode only the diagonal elements, i.e.

$$F_{ab} = \frac{f_{\rm sky}}{6}\sum_{\ell_i} I_{\ell_1\ell_2\ell_3}\frac{b^{(a)}_{\ell_1\ell_2\ell_3}b^{(b)}_{\ell_1\ell_2\ell_3}}{C_{\ell_1}C_{\ell_2}C_{\ell_3}}, \tag{291}$$

where

$$I_{\ell_1\ell_2\ell_3} = \sum_{m_i}(\mathcal{G}^{m_1m_2m_3}_{\ell_1\ell_2\ell_3})^2 = \frac{(2\ell_1+1)(2\ell_2+1)(2\ell_3+1)}{4\pi} \tag{292}$$

Because of symmetries we can write:

$$F_{ab} = f_{\rm sky}\sum_{\ell_1\geq\ell_2\geq\ell_3} I_{\ell_1\ell_2\ell_3}\frac{b^{(a)}_{\ell_1\ell_2\ell_3}b^{(b)}_{\ell_1\ell_2\ell_3}}{C_{\ell_1}C_{\ell_2}C_{\ell_3}\Delta_{\ell_1\ell_2\ell_3}}, \tag{293}$$

where $\Delta_{\ell_1\ell_2\ell_3} = 1, 2$ and 6 when all ℓ_i are different, two are the same, or all three are identical respectively. We can use the Fisher Matrix to make predictions about the detectibility of given non-Gaussian amplitude $f^{(a)}_{\rm NL}$. We show several predictions in Fig 6 for local, equilateral and orthogonal spectra where we assume no noise and $f_{\rm sky} = 1$. The main take-away here is that even for the most optimistic case, it will be challenging (and downright impossible) to detect $f_{\rm NL} = 1$ for these 3 shapes that have been identified as key markers of inflaton dynamics.

Lets us define the following functions (KSW 2003):

$$\alpha_\ell(\chi) = \frac{2}{\pi}\int k^2 dk\,\Theta^{T,\mathcal{R}}_\ell(k) j_\ell(k\chi), \tag{294}$$

$$\beta_\ell(\chi) = \frac{2}{\pi}\int k^2 dk\,P_{\mathcal{R}}(k)\Theta^{T,\mathcal{R}}_\ell(k) j_\ell(k\chi). \tag{295}$$

We can then write the reduced local bispectrum as:

$$b^{\rm local}_{\ell_1\ell_2\ell_3} = f^{\rm local}_{\rm NL} 2\int d\chi\chi^2\left[\beta_{\ell_1}(\chi)\beta_{\ell_2}(\chi)\alpha_{\ell_3}(\chi) + 2\text{ perms}\right]. \tag{296}$$

The above expression would suffice in case we are interested in performing a Fisher forecast. For the estimator we need $\mathcal{S}_{\rm local}$. The observed bispectrum will also pick up an experimental window function w_ℓ that captures the effect of the beam and pixilation:

$$b^{\rm local}_{\ell_1\ell_2\ell_3} \to f^{\rm local}_{\rm NL} 2\int d\chi\chi^2\left[\beta_{\ell_1}(\chi)\beta_{\ell_2}(\chi)\alpha_{\ell_3}(\chi) + 2\text{ perms}\right] w_{\ell_1}w_{\ell_2}w_{\ell_3}. \tag{297}$$

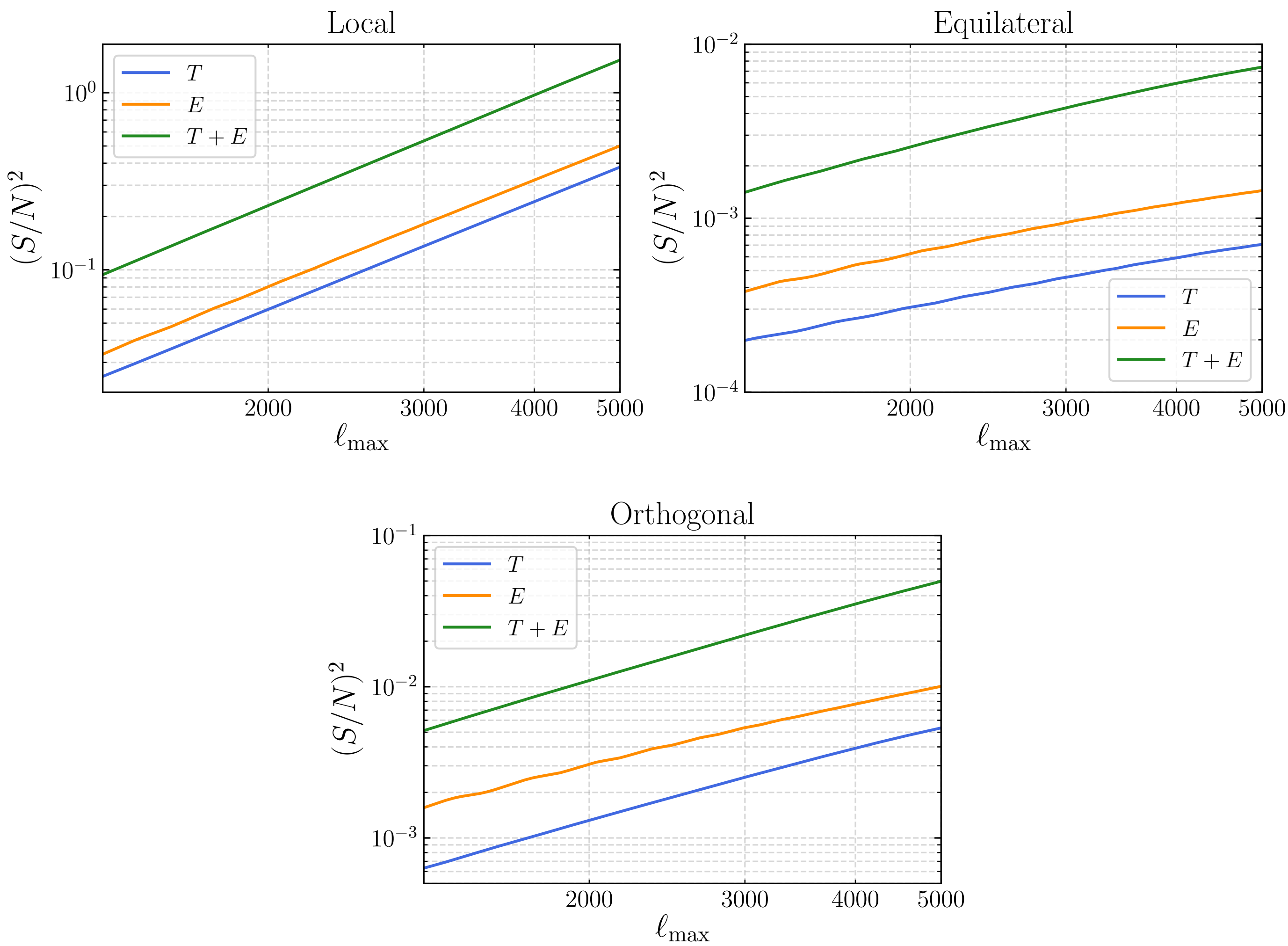

Figure 6: Numerical solution of the bispectrum $(S/N)^2$ derived from equation Eq. (293) in the range $\ell_{\rm d} \leq \ell_{\rm max} \leq 5000$, where $\ell_{\rm d} \simeq 1300$. *Upper left panel:* $(S/N)^2$ for the local shape for temperature, E-mode polarization and in combination. *Upper right and lower panel*: $(S/N)^2$ for the equilateral and orthogonal shapes respectively. Curves shown are for $f_{\rm NL} = 1$. Figure adopted from Kalaja et al 2021.

With this expression, we can write down $\mathcal{S}_{\rm local}$ (KSW 2003)

$$\mathcal{S}_{\rm local} = \int \chi^2 d\chi \int d^2\hat{\mathbf{n}} \left[A(\hat{n},\chi)B^2(\hat{n},\chi) - 2B(\hat{n},\chi)\langle A(\hat{n},\chi)B(\hat{n},\chi)\rangle_{\rm MC} - A(\hat{n},\chi)\langle B^2(\hat{n},\chi)\rangle_{\rm MC}\right], \tag{298}$$

where we have defined:

$$A(\hat{n},\chi) \equiv \sum_\ell w_\ell \alpha_\ell(\chi) e_\ell(\hat{n}) = \sum_{\ell m} w_\ell \alpha_\ell(\chi)(C^{-1}a)_{\ell m} Y_{\ell m}(\hat{n}), \tag{299}$$

$$B(\hat{n},\chi) \equiv \sum_\ell w_\ell \beta_\ell(\chi) e_\ell(\hat{n}) = \sum_{\ell m} w_\ell \beta\ell(\chi)(C^{-1}a)_{\ell m} Y_{\ell m}(\hat{n}). \tag{300}$$

We set $f_{\rm NL}^{\rm local} = 1$, i.e. our estimator is matched to obtain an optimal estimate at this value.

Similarly, using the equilateral template, we can derive $\mathcal{S}_{\rm equil}$:

$$\mathcal{S}_{\rm equil} = -3\mathcal{S}_{\rm local} + 6\int \chi^2 d\chi \int d^2\hat{n} \left[BCD - B\langle CD\rangle_{\rm MC} - C\langle BD\rangle_{\rm MC} - D\langle BC\rangle_{\rm MC} - \frac{1}{3}(D^3 - 3D\langle D^2\rangle_{\rm MC})\right] \tag{301}$$

where I dropped $(\hat{n},\chi)$ for brevity. Here we have defined in addition:

$$C(\hat{n},\chi) \equiv \sum_\ell w_\ell \gamma\ell(\chi) e_\ell(\hat{n}) = \sum_{\ell m} w_\ell \gamma_\ell(\chi)(C^{-1}a)_{\ell m} Y_{\ell m}(\hat{n}), \tag{302}$$

$$D(\hat{n},\chi) \equiv \sum_\ell w_\ell \delta_\ell(\chi) e_\ell(\hat{n}) = \sum_{\ell m} w_\ell \delta\ell(\chi)(C^{-1}a)_{\ell m} Y_{\ell m}(\hat{n}), \tag{303}$$

with

$$\gamma_\ell(\chi) = \frac{2}{\pi}\int k^2 dk P_{\mathcal{R}}^{1/3}(k)\Theta_\ell^{T,\mathcal{R}}(k) j_\ell(k\chi), \tag{304}$$

$$\delta_\ell(\chi) = \frac{2}{\pi}\int k^2 dk P_{\mathcal{R}}^{2/3}(k)\Theta_\ell^{T,\mathcal{R}}(k) j_\ell(k\chi). \tag{305}$$

Note that for polarization, we need to replace the transfer functions with those for E. The above decomposition allows us to build an efficient estimator for most factorized shapes by appropriately using products of the functions α, β, δ and γ. If a shape is not factorized, we can use the binned or modal estimators.

The full estimator would also include E- and B-mode polarisation (Duivenvoorden, Meerburg, Freese 2020):

$$\hat{f}^a_{\rm NL} = \frac{1}{6\mathcal{I}^a_0}\sum_{{\rm all}\,\ell,m}\sum_{{\rm all}X}(B_a)^{\ell_1\ell_2\ell_3}_{m_1m_2m_3,X_1X_2X_3} \times \left\{ \left[(C^{-1}a)^{X_1}_{\ell_1m_1}(C^{-1}a)^{X_2}_{\ell_2m_2}(C^{-1}a)^{X_3}_{\ell_3m_3}\right] - \left[(C^{-1})^{X_1X_2}_{\ell_1m_1\ell_2m_2}(C^{-1}a)^{X_3}_{\ell_3m_3} + {\rm cyclic}\right]\right\} \tag{306}$$

Here $X \in \{T, E, B\}$ and

$$(C^{-1}a)^X_{\ell m} = \sum_{X'}\sum_{\ell',m'}(C^{-1})^{XX'}_{\ell m\ell' m'} a_{X',\ell' m'}. \tag{307}$$

The inverse C^{-1} is taken from:

$$C_{\ell m\ell' m'} \equiv \begin{pmatrix} C_{TT} & C_{TE} & C_{TB} \\ C_{ET} & C_{EE} & C_{EB} \\ C_{BT} & C_{BE} & C_{BB} \end{pmatrix}_{\ell m\ell' m'}. \tag{308}$$

$\mathcal{I}^a_0$ is the 'normalisation' and is simply the (data independent) Fisher information on shape (a), i.e. F_a:

$$\mathcal{I}^a_0 = \frac{1}{6}\sum_{{\rm all}\,\ell,m}\sum_{{\rm all}X}(B_a)^{\ell_1\ell_2\ell_3}_{m_1m_2m_3,X_1X_2X_3} \times \left[(C^{-1})^{X_1X_4}_{\ell_1m_1\ell_4m_4}(C^{-1})^{X_2X_5}_{\ell_2m_2\ell_5m_5}(C^{-1})^{X_3X_6}_{\ell_3m_3\ell_6m_6}\right] \times (B^*_a)^{\ell_4\ell_5\ell_6}_{m_4m_5m_6,X_4X_5X_6}. \tag{309}$$

The above allows us to constrain (or detect) any primordial bispectrum sources by scalars $\mathcal{R}$. For mixed primordial non-Gaussianities (which we have not discussed but do exist) that couple γ_{ij} to $\mathcal{R}$, building an efficient and factorized KSW like estimators is a little more involved due to the helicity structure of the gravitational waves. This explicitly coupled modes and requires a small number of additional terms (see Duivenvoorden, Meerburg and Freese). Alternatively one might use the modal or binned estimators instead.

Next, lets discuss the current constraints. Unlike constraints and measurements of MOST cosmological parameters, constraining primordial non-Gaussianity is a little different. We have derived an optimal estimator based on maximising the PDF for a given amplitude. We assumed all other cosmological parameters hidden in $P_{\mathcal{R}}$ (in fact we typically set $n_s = 1$). and Θ_ℓ are fixed. This is purely driven by computational limitations; it is impossible to 'sample' bispectra at the moment. Also, since we have not detected any deviations from non-Gaussianity, sampling the PDF will not really help; most parameters are already very well determined from the CMB power spectrum. In summary, estimating $\hat{f}_{NL}$ is a one-shot effort and the error on the estimate is derived from applying the estimator on a bunch of null maps. We also assume that the error on $f_{\rm NL}$ is hardly affected by the presence of non-Gaussianity in the data, i.e. we assume that he covariance is Gaussian. So far this has been a reasonable assumption, but it turns out that this will no longer suffice for future measurements of the CMB bispectrum (it never holds except on very large scale when measuring the bispectrum of large scale structure).

To get a feeling for whet we are looking for, we generated some maps containing primordial non-Gaussianity of the local type and show these for different values of $f_{\rm NL}^{\rm local}$ in Fig. 7. The key take-away here is that we are looking for a needle in a haystack. First constraints were derived from COBE data (Komatsu thesis 2002). Significant improvements

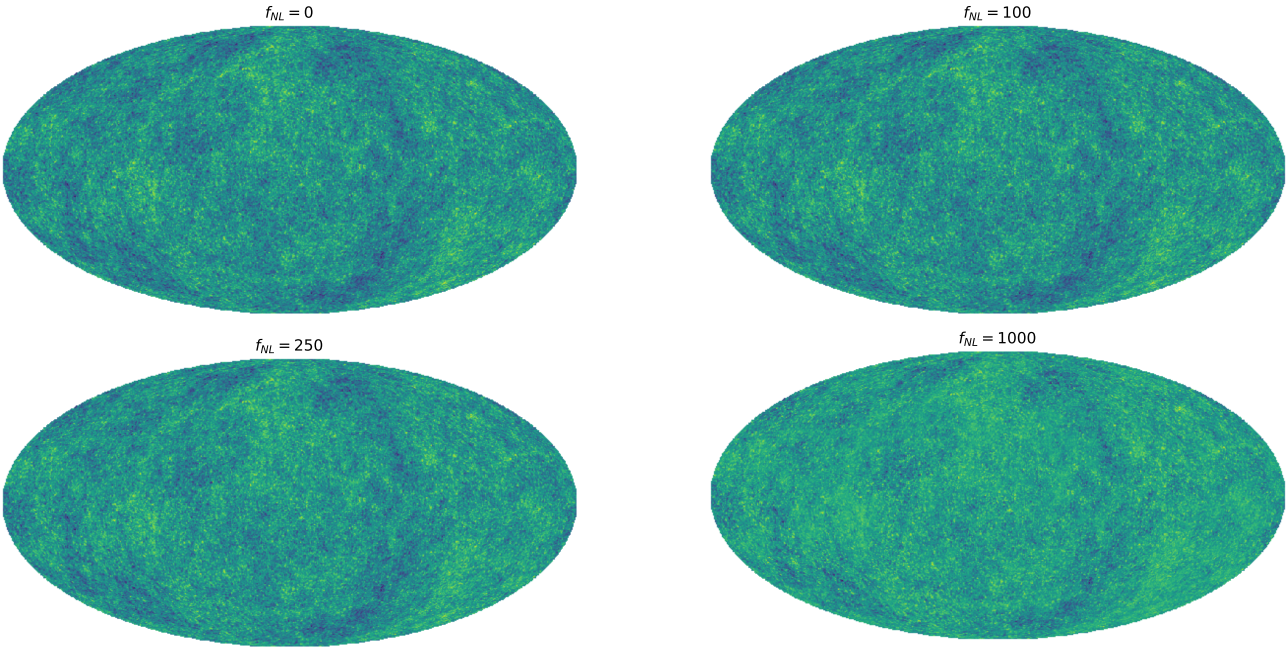

Figure 7: Simulated CMB maps with different levels of primordial non-Gaussianity of the local type. Only for values $f_{\rm NL}^{\rm local}$ can we see differences in the map by eye. The obvious reason is that since the fluctuations in the map are of order $\sqrt{A_s}$, the non-Gaussian amplitude is suppressed by this factor compared to the Gaussian fluctuations. Hence only once we offset this by considering an $f_{\rm NL}^{\rm local} = \mathcal{O}(1/\sqrt{A_s})$ do we start seeing a significant difference. Note that this is actually a regime in which are theoretical computations would no longer make any sense. We have assumed weak non-Gaussianity throughout. If primordial non-Gaussianity would be indeed of this order, then all higher order spectra would become $\mathcal{O}(1)$.

were derived from WMAP and more recently Planck data[20]:

$$f_{\rm NL}^{\rm local} = -0.9 \pm 5.1 \tag{310}$$

$$f_{\rm NL}^{\rm equil} = -26 \pm 4 \tag{311}$$

$$f_{\rm NL}^{\rm ortho} = -38 \pm 24 \tag{312}$$

These are currently the tightest constraints on primordial non-Gaussianity. Using the modal and binned estimators a large number of other primordial bispectra have been constrained, and none of these passes the threshold of a detection.

[20]There is one caveeat. The constraint on $f_{\rm NL}^{\rm ortho}$ was derived using a template that captures the shape but does not have the correct soft limit. This gives constraints that are somewhat better than the one you should get if you considered the template of Eq. (279).