Prepared for submission to JHEP

# EFTs and an Introduction to Inflation

**C.P. Burgess**

*Department of Physics & Astronomy, McMaster University*
*and Perimeter Institute for Theoretical Physics*

Abstract: PRELIMINARY: INCOMPLETE FINAL SECTIONS AND REFERENCES
These notes present an introduction to inflationary cosmology with an emphasis on some of the ways effective field theories are used in its analysis. Based on lectures prepared for the Les Houches Summer School *Effective Field Theory in Particle Physics and Cosmology*, July 2017. Parts of these lectures draw on my earlier notes *Lectures on Cosmic Inflation and its Potential Stringy Realizations* [1] as well as *Quantum Gravity in everyday life: General Relativity as an Effective Field Theory* [2] and *Who You Gonna Call? Runaway Ghosts, Higher Derivatives and Time-Dependence in EFTs* [3].

## Contents

---

These lectures are meant to provide a brief overview of two topics: the standard (Hot Big Bang, or ΛCDM) model of cosmology and the inflationary universe that presently provides our best understanding of the standard cosmology's peculiar initial conditions. There are several goals to this presentation: the first of which is to provide a particle-physics audience with some of the tools required by later lecturers in this school. After all, cosmology has become a mainstream topic within particle physics, largely because cosmology provides several of the main pieces of observational evidence for the incompleteness of the Standard Model of particle physics.

A second goal of these lectures is to touch on the important role played in cosmology by many of the same methods of effective field theory (EFT) used elsewhere in physics. This second goal is particularly important for the cosmology of the very early universe (such as inflationary or 'bouncing' models) for which a central claim is that quantum fluctuations provide an explanation of the properties of primordial fluctuations presently found writ large across the sky. If true, this claim would imply not only that quantum gravity effects are observable; the claim is that their imprint has already been observed cosmologically. Such claims sharpen the need to clarify what parameters control the size of quantum effects in gravity, and along the way more generally to identify the domain of validity of semi-classical methods is cosmology.

In practice the lectures are divided into two parts: homogeneous, isotropic cosmologies and the fluctuations about them. The first part provides a very brief description of the classic homogeneous and isotropic cosmological models usually encountered in introductory cosmology courses. One goal of this section is to highlight both the great success these models have describing the Universe as we find it around us. The second goal is to describe the peculiar initial conditions that are required by this observational success. This section then highlights how these puzzling initial conditions suggest the Universe once underwent an earlier epoch of accelerated expansion. It closes by describing several simple and representative single-field inflationary models that have been proposed to provide this earlier accelerated epoch.

The second part of the lectures repeats the same picture, but now for fluctuations about both standard and inflationary cosmologies. This section starts by describing

the very successful picture of structure formation within the standard ΛCDM model, in which both fluctuations in the cosmic microwave background (CMB) and the distribution of galaxies is attributed to the amplification by gravity of a simple primordial spectrum of small fluctuations. Again the success of standard cosmology proves to rely on a specific choice for the initial spectrum of primordial fluctuations, and again the required initial spectrum can be understood as being produced by quantum fluctuations if there were an earlier epoch of accelerated expansion. Accelerated expansion plays double duty: potentially both explaining the initial conditions of the background homogeneous Universe and of the primordial spectrum of fluctuations within it.

Because of the important role played by gravitating quantum fluctuations, EFT methods are central to assessing the domain of validity of the entire picture. Consequently the the third section of these notes summarizes several of the ways they do so, and how their application can differ in cosmology from those encountered elsewhere in particle physics. This starts by extending standard power-counting arguments to identify the small parameters that control the underlying semiclassical expansion implicitly used in essentially all cosmological models. In passing we comment on why control over the semiclassical expansion tends to favour inflationary models over their alternatives (such as bouncing cosmologies).[1] Other EFT topics discussed include several new issues of principle to do with how to define EFTs in explicitly time-dependent situations, and quantifying the robustness of inflationary predictions to any peculiarities of unknown higher-energy physics. This section closes with a short description of the practical EFT of fluctuations in single-field inflationary models used to identify potential observational signals in as model-independent way as possible.

# 1 Cosmology: Background

This section summarizes the standard discussion of background cosmology, both for ΛCMD models and their inflationary precursors.

## 1.1 Standard ΛCDM cosmology

The starting point is the standard cosmology of the expanding Universe revealed to us by a astronomical observations.

[1] Of course, although this explains the current preference amongst cosmologists for inflationary models, it does not mean that Nature prefers them. Rather, these EFT arguments just set the bar to which formulations of alternative proposals should also aspire to achieve equal credence.

### 1.1.1 FRW geometries

Cosmology became a science once Einstein's discovery of General Relativity related the observed distribution of stress-energy to the measurable geometry of space-time. This implies the geometry of the Universe as a whole can be tied to the overall distribution of matter at the largest scales. These days it is an experimental fact that the stress-energy of the Universe appears to be very homogeneous and isotropic on the largest scales visible. One piece of evidence to this effect is the very small — one part in $10^5$ — temperature fluctuations of the CMB (more about which later).

On such large scales the geometry of space-time might also be expected to be homogeneous and isotropic, and the most general such a geometry in 3+1 dimensions is described by the Friedmann-Robertson-Walker (FRW) metric. The line-element for this metric can be written as[2]

$$\begin{aligned} \mathrm{d}s^2 = g_{\mu\nu}\,\mathrm{d}x^\mu \mathrm{d}x^\nu &= -\mathrm{d}t^2 + a^2(t)\,\left[\frac{\mathrm{d}r^2}{1-\kappa r^2/R_0^2} + r^2\,\mathrm{d}\theta^2 + r^2\sin^2\theta\,\mathrm{d}\phi^2\right] \\ &= -\mathrm{d}t^2 + a^2(t)\,\left[\mathrm{d}\ell^2 + r^2(\ell)\,\mathrm{d}\theta^2 + r^2(\ell)\sin^2\theta\,\mathrm{d}\phi^2\right]\,, \end{aligned} \tag{1.1}$$

where $R_0$ is a constant with dimension length and $\kappa$ can take one of the following three values: $\kappa = 1, 0, -1$. The coordinate $\ell$ is related to $r$ by $\mathrm{d}\ell = \mathrm{d}r/(1-\kappa r^2/R_0^2)^{1/2}$, and so

$$r(\ell) = \begin{cases} R_0\,\sin(\ell/R_0) & \text{if} \quad \kappa = +1 \\ \ell & \text{if} \quad \kappa = 0 \\ R_0\,\sinh(\ell/R_0) & \text{if} \quad \kappa = -1 \end{cases}\,. \tag{1.2}$$

The quantity $a(t)R_0$ represents the radius of curvature of the spatial slices at fixed $t$, which are 3-spheres when $\kappa = 0$; 3-hyperbolae for $\kappa = -1$ and are flat for $\kappa = 0$. It is conventional to scale $R_0$ out of the metric by re-scaling the coordinates $\ell \to R_0\,\ell$ and $r \to R_0\,r$ while at the same time rescaling $a(t) \to a(t)/R_0$. This redefinition makes $r$ and $\ell$ dimensionless while giving $a(t)$ units of length, and it is often useful to choose cosmological units for which $a(t_0) = 1$ for some $t_0$ (such as at present). The case $\kappa = 0$ turns out to be of particular interest because all current evidence (coming, for instance, from the measured properties of the CMB) indicates that the spatial slices in the Universe are consistent with being flat.

Trajectories along which only $t$ varies are time-like geodesics of this metric and represent the motion of a natural set of static 'co-moving' observers. The co-moving

[2]For those rusty on what a metric means and perhaps needing a refresher course on General Relativity using the same conventions as those used here, feel free to refresh you memory with *General Relativity: The Notes* at `http/www.physics.mcmaster.ca/~cburgess/Notes/GRNotes/pdf`.

distance, $\Delta\ell$, between two such observers at a fixed time $t$ is related to their physical distance — as measured by the metric (1.1) — by

$$D(\Delta\ell, t) = \Delta\ell\, a(t)\,, \tag{1.3}$$

so the 'scale-factor' $a(t)$ describes the common time-evolution of spatial scales. So long as $a(t)$ is monotonic one can use $t$ or $a$ interchangeably as measures of the passage of time.

The trajectories of photons play a special role in cosmology since until very recently they brought us all of our information about the universe at large. Since they move at the speed of light their trajectories satisfy $\mathrm{d}s^2 = 0$ and so

$$g_{\mu\nu}\left(\frac{\mathrm{d}x^\nu}{\mathrm{d}s}\right)\left(\frac{\mathrm{d}x^\nu}{\mathrm{d}s}\right) = 0\,, \tag{1.4}$$

which for radial motion specializes to $\mathrm{d}t/\mathrm{d}s = \pm a(t)(\mathrm{d}\ell/\mathrm{d}s)$. Choosing coordinates that place us at the origin means all photons sent to us move along a radial trajectory.

A photon arriving at $t = 0$ from a galaxy situated at fixed co-moving position $\ell = L$ must have departed at time $t = -T$ where

$$L = \int_0^T \frac{\mathrm{d}t}{a(t)}\,. \tag{1.5}$$

Since the universe expands by an amount $a_0/a$ in this time (where $a_0 = a(0)$ is the present-day scale factor and $a = a(-T)$ is its value when the light was emitted), the redshift, $z$, of the light is given by $z := (\lambda_{\rm obs} - \lambda_{\rm em})/\lambda_{\rm em}$, with $\lambda_{\rm obs}/\lambda_{\rm em} = a_0/a$. Consequently $z$ and $a$ are related by

$$1 + z = \frac{a_0}{a}\,. \tag{1.6}$$

This very usefully ties the Universal expansion to the more easily measured redshift of distant objects.[3]

For later purposes, it is worth introducing another useful time coordinate when discussing the evolution of light rays in FRW geometries. Defining 'conformal time', $\tau$, by

$$\tau = \int \frac{\mathrm{d}t}{a(t)}\,, \tag{1.7}$$

[3]In practice the redshift of any particular object depends also on its 'peculiar' motion relative to the co-moving observers, but in practice this is negligible compared with the cosmic redshift for all but relatively nearby galaxies.

allows the metric (1.1) to be written

$$\mathrm{d}s^2 = a^2(\tau)\,\left[-\mathrm{d}\tau^2 + \mathrm{d}\ell^2 + r^2(\ell)\,\mathrm{d}\theta^2 + r^2(\ell)\sin^2\theta\,\mathrm{d}\phi^2\right]\,. \tag{1.8}$$

The utility of this coordinate system is that the scale-factor $a(\tau)$ completely drops out of the evolution of photons, which simplifies the identification of many of the causal properties of the spacetime (*i.e.* identifying which events can communicate with each other by exchanging photons).

### 1.1.2 Implications of Einstein's equations

So far so good, but the story so far is largely just descriptive. The FRW metric, with $a(t)$ specified, says much about how particles move over cosmological distances. But we also need to know how to relate $a(t)$ to the Universe's stress-energy content. This connection is made using Einstein's equations,[4]

$$R_{\mu\nu} - \frac{1}{2}\,R\,g_{\mu\nu} + 8\pi G\,T_{\mu\nu} = 0\,, \tag{1.9}$$

where $G$ is Newton's constant of universal gravitation, $R_{\mu\nu} = R^{\alpha}{}_{\mu\alpha\nu}$ is the geometry's Ricci tensor (where $R^{\alpha}{}_{\mu\beta\nu}$ is its Riemann tensor) and $R = g^{\mu\nu}R_{\mu\nu}$.

The twin requirements of homogeneity and isotropy dictate that the most general form for the Universe's stress-energy tensor, $T_{\mu\nu}$, is that of a perfect fluid,

$$T_{\mu\nu} = p\,g_{\mu\nu} + (p+\rho)\,U_\mu U_\nu\,, \tag{1.10}$$

where $p$ and $\rho$ are respectively the fluid's pressure and energy density, while $U^\mu\partial_\mu = \partial_t$ (or, equivalently, $U_\mu\,\mathrm{d}x^\mu = -\mathrm{d}t$) is the 4-velocity of the co-moving observers.

Specialized to the metric (1.1) the Einstein equations boil down to the following two independent equations:

$$H^2 + \frac{\kappa}{a^2} = \frac{8\pi G}{3}\,\rho = \frac{\rho}{3M_p^2} \qquad \text{(Friedmann equation)} \tag{1.11}$$

and

$$\dot\rho + 3H(p+\rho) = 0 \qquad \text{(energy conservation)} \tag{1.12}$$

where over-dots denote differentiation with respect to $t$ and the Hubble function is defined by $H = \dot a/a$. The last equality in eq. (1.11) also defines the 'reduced' Planck

[4] Besides using metric signature $(-+++)$, unless explicitly stated otherwise I also use units with $\hbar = c = k_B = 1$, and follow Weinberg's curvature conventions [4] (which differ from the popular MTW conventions [5] only by an overall sign in the definition of the Riemann curvature, $R^{\mu}{}_{\nu\lambda\rho}$).

mass: $M_p^2 = (8\pi G)^{-1} \simeq 10^{18}$ GeV. Differentiating (1.11) and using (1.12) gives a useful formula for the cosmic acceleration

$$\frac{\ddot{a}}{a} = -\frac{1}{6M_p^2}(\rho + 3p)\,. \tag{1.13}$$

Mathematically speaking, finding the evolution of the universe as a function of time requires the integration of eqs. (1.11) and (1.12), but in themselves these two equations are inadequate to determine the evolution of the three unknown functions, $a(t)$, $\rho(t)$ and $p(t)$. Another condition is required in order to make the problem well-posed. The missing condition is furnished by the equation of state for the matter in question, which for the present purposes we take to be an expression for the pressure as a function of energy density, $p = p(\rho)$. In particular, the equations of state of interest in ΛCDM cosmology have the general form

$$p = w\,\rho\,, \tag{1.14}$$

where $w$ is a $t$-independent constant.

The first step in solving for $a(t)$ is to determine how $p$ and $\rho$ depend on $a$, since this is dictated by energy conservation. Using eq. (1.14) in (1.12) allows it to be integrated to obtain

$$\rho = \rho_0 \left(\frac{a_0}{a}\right)^{\sigma} \qquad \text{with} \quad \sigma = 3(1+w)\,. \tag{1.15}$$

Eq. (1.14) implies the pressure satisfies an identical dependence on $a$. Similarly using eq. (1.15) to eliminate $\rho$ from (1.11) leads to the following differential equation for $a(t)$:

$$\dot{a}^2 + \kappa = \frac{8\pi G\rho_0 a_0^2}{3}\left(\frac{a_0}{a}\right)^{\sigma-2}\,. \tag{1.16}$$

When $\kappa = 0$ this equation is easily integrated to give

$$a(t) = a_0\left(\frac{t}{t_0}\right)^{\alpha} \qquad \text{with} \quad \alpha = \frac{2}{\sigma} = \frac{2}{3(1+w)}\,. \tag{1.17}$$

### 1.1.3 Equations of state

In the ΛCDM model of cosmology the total energy density is regarded as the sum of several components, each of which separately satisfies one of the following three basic equations of state.

#### Nonrelativistic matter

An ideal gas of non-relativistic particles in thermal equilibrium has a pressure and energy density given by

$$p = n\,T \qquad \text{and} \qquad \rho = n\,m + \frac{n\,T}{\gamma - 1}\,, \tag{1.18}$$

where $n$ is the number of particles per unit volume, $m$ is the particle's rest mass and $\gamma = c_p/c_v$ is its ratio of specific heats, with $\gamma = 5/3$ for a gas of monatomic atoms. For non-relativistic particles the total number of particles is usually also conserved,[5] which implies that

$$\frac{\mathrm{d}}{\mathrm{d}t}\Big[n\,a^3\Big] = 0\,. \tag{1.19}$$

Since $m \gg T$ (or else the atoms would be relativistic) the equation of state for this gas may be taken to be

$$\frac{p}{\rho} \sim \frac{T}{m} \ll 1 \qquad \text{and so} \qquad w \simeq 0\,. \tag{1.20}$$

Since $w \simeq 0$ energy conservation implies $\sigma = 3(1+w) \simeq 3$ and so $\rho\,a^3$ is a constant. This is appropriate for nonrelativistic matter for which the energy density is dominated by the particle rest-masses, $\rho \simeq n\,m$, because in this case energy conservation is equivalent to conservation of particle number, which (1.19) states implies $n \propto a^{-3}$.

Finally, whenever the total energy density is dominated by non-relativistic matter we know $w = 0$ also implies $\alpha = 2/\sigma = 2/3$ and so if $\kappa = 0$ then the universal scale factor expands like $a \propto t^{2/3}$.

**Radiation**

Thermal equilibrium dictates that a gas of relativistic particles (like photons) must have an energy density and pressure given by

$$\rho = a_B\,T^4 \qquad \text{and} \qquad p = \frac{1}{3}\,a_B\,T^4\,, \tag{1.21}$$

where $a_B = \pi^2/15 = 0.6580$ is the Stefan-Boltzmann constant (in units where $k_B = c = \hbar = 1$) and $T$ is the temperature. Together, these ensure that $\rho$ and $p$ satisfy the equation of state

$$p = \frac{1}{3}\,\rho \qquad \text{and so} \qquad w = \frac{1}{3}\,. \tag{1.22}$$

Eq. (1.22) also applies to any other particle whose temperature dominates its rest mass, and so in particular applies to neutrinos for most of the Universe's history.

[5] If their number happens not to be both conserved and constrained to be nonzero, then once the temperature becomes low enough ($T \lesssim m$) for nonrelativistic kinematics to apply their density becomes quite small if they remain in thermal equilibrium. This is due to the Boltzmann suppression, $n \propto e^{-m/T}$, that arises because at these temperatures the annihilation of particles and antiparticles is not compensated by their pair-production, due to there being insufficient thermal energy.

Since $w = 1/3$ it follows that $\sigma = 3(1+w) = 4$ and so $\rho \propto a^{-4}$. This has a simple physical interpretation for a gas of noninteracting photons, since for these the total number of photons is fixed and so $n_\gamma \propto a^{-3}$. But each photon energy is inversely proportional to its wavelength and so also redshifts like $1/a$ as the universe expands, leading to $\rho_\gamma \propto a^{-4}$.

Whenever radiation dominates the total energy density then $w = 1/3$ implies $\alpha = 2/\sigma = 1/2$, and so if $\kappa = 0$ then $a(t) \propto t^{1/2}$.

**The vacuum**

If the vacuum is Lorentz invariant, as the success of special relativity seems to indicate, then its stress energy must satisfy $T_{\mu\nu} \propto g_{\mu\nu}$. This implies the vacuum pressure must satisfy the only possible Lorentz-invariant equation of state:

$$p = -\rho \qquad \text{and so} \qquad w = -1\,. \tag{1.23}$$

Because $w = -1$ we have $\sigma = 3(1+w) = 0$ and so energy conservation implies that $\rho$ is a constant, independent of $a$ or $t$. This kind of constant energy density is often called, for historical reasons, a cosmological constant. Although counter-intuitive, constant energy density can be consistent with energy conservation in an expanding Universe. This is because (1.12) implies the total energy satisfies $\mathrm{d}(\rho\, a^3)/\mathrm{d}t = -p\, \mathrm{d}(a^3)/\mathrm{d}t$. This shows that the equation of state (1.23) ensures the pressure does precisely the amount of work required to produce the change in total energy required by having constant energy density.

When the vacuum dominates the energy density then $\alpha = 2/\sigma \to \infty$, which shows that the power-law solutions, $a \propto t^\alpha$, are not appropriate. Returning directly to the Friedmann equation, eq. (1.11), shows (when $\kappa = 0$) that $H = \dot{a}/a$ is constant and so the solutions are exponentials: $a \propto \exp[\pm H(t - t_0)]$. Notice that (1.23) implies $\rho + 3p$ is negative if $\rho$ is positive. This furnishes an explicit example of an equation of state for which the universal acceleration, $\ddot{a}/a = -\frac{4}{3}\pi G(\rho + 3p)$, can be positive.

### 1.1.4 Universal energy content

At present there is direct observational evidence that the universe contains at least 4 independent types of matter, whose properties are now briefly summarized.

**Radiation**

The universe is known to be awash with photons, and is also believed contain similar numbers of neutrinos (that until very recently[6] could also be considered to be radiation).

*Cosmic Photons:*

The most numerous type of photons found at present in the Universe are the photons in the cosmic microwave background (CMB). These are distributed thermally in energy with a temperature that is measured today to be $T_{\gamma 0} = 2.725$ K. The present number density of these CMB photons is determined by their temperature to be

$$n_{\gamma 0} = 4.11 \times 10^8 \text{ m}^{-3}\,, \tag{1.24}$$

which turns out to be much higher than the number density of ordinary atoms. Their present energy density (also determined by their temperature) is

$$\rho_{\gamma 0} = 0.261 \text{ MeV m}^{-3} \qquad \text{or} \qquad \Omega_{\gamma 0} = 5.0 \times 10^{-5}\,, \tag{1.25}$$

where $\Omega_{\gamma 0} := \rho_{\gamma 0}/\rho_{c0}$ defines the fraction of the total energy density (also the 'critical' density, $\rho_{c0} \simeq 5200$ MeV $^{-3} \simeq 10^{-29}$ g cm$^{-3}$) currently residing in CMB photons.

*Relict Neutrinos:*

It is believed on theoretical grounds that there are also as many cosmic relict neutrinos as there are CMB photons running around the universe, although these neutrinos have never been detected. They are expected to have been relativistic until relatively recently in cosmic history, and to be thermally distributed. The neutrinos are expected to have a slightly lower temperature, $T_{\nu 0} = 1.9$ K, and are fermions and so have a slightly different energy-density/temperature relation than do neutrinos.

Their contribution to the present-day cosmological energy budget is not negligible, and if they were massless would be predicted to be

$$\rho_{\nu 0} = 0.18 \text{ MeV m}^{-3} \qquad \text{or} \qquad \Omega_{\nu 0} = 3.4 \times 10^{-5}\,, \tag{1.26}$$

leading to a total radiation density, $\Omega_{R0} = \Omega_{\gamma 0} + \Omega_{\nu 0}$, of size

$$\rho_{R0} = 0.44 \text{ MeV m}^{-3} \qquad \text{or} \qquad \Omega_{r0} = 8.4 \times 10^{-5}\,. \tag{1.27}$$

[6] Although neutrino masses play an important role in some things (like the formation of galaxies and other structure), I lump them here with radiation because for most of what follows the fact that they very recently likely became nonrelativistic does not matter.

### Baryons

The main constituents of matter we see around us are atoms, made up of protons, neutrons and electrons, and these are predominantly non-relativistic at the present epoch. Furthermore the total abundance of electrons is very likely precisely equal to that of protons in order to ensure that the universe carries no net charge.

Since protons and neutrons are about 1840 times more massive than electrons, the energy density in ordinary non-relativistic particles is likely to be well approximated by the total energy in protons and neutrons: the total energy in baryons. It turns out it is possible to determine the total number of baryons in the universe (regardless of whether or not they are presently visible), in several independent ways.

One way to determine the baryon density uses measurements of the properties of the CMB, whose understanding depends on things like the speed of sound or on reaction rates – and so also on the density – for the Hydrogen gas from which the CMB photons last scattered. Another way uses the success of the predictions for the abundances of light elements as nuclei formed during the very early universe, which depends on nuclear reaction rates – again proportional to the total nucleon density.

These two kinds of inferences are consistent with each other and indicate the total energy density in baryons is

$$\rho_{B0} = 210 \text{ MeV m}^{-3} \qquad \text{or} \qquad \Omega_{B0} = 0.04\,. \tag{1.28}$$

For purposes of comparison, this is about ten times larger than the amount of *luminous* matter, found using the luminosity density for galaxies, $nL = 2 \times 10^8\ L_\odot\ \text{Mpc}^{-3}$, together with the best estimates of the average mass-to-luminosity ratio of for galactic matter: $M/L \simeq 4M_\odot/L_\odot$.

It should be emphasized that although there is more energy in baryons than in CMB photons, the *number density* of baryons is much smaller, since

$$n_{B0} = \frac{210 \text{ MeV m}^{-3}}{940 \text{ MeV}} = 0.22 \text{ m}^{-3} = 5 \times 10^{-10}\, n_{\gamma 0}\,. \tag{1.29}$$

### Dark Matter

There several lines of evidence pointing to the large-scale presence of another form of non-relativistic matter besides baryons, carrying much more energy than do the baryons. Part of the evidence for this so-called *Dark Matter* comes from a variety of independent ways of measuring of the total amount of gravitating mass in galaxies and in clusters of galaxies. The rotation rates of galaxies indicate that there is considerably

more gravitating mass present than would be inferred by counting the luminous matter which can be seen. A similar result holds for the total mass in galaxy clusters, as estimated from the motions of their constituent galaxies, from the temperature of their hot inter-galactic gas and from the amounts of gravitational lensing which they produce. Furthermore, whatever it is this matter should be non-relativistic since it takes part in the gravitational collapse which gives rise to galaxies and their clusters. (Relativistic matter tends not to cluster in this way, as we see in later sections.)

All of these estimates appear to be consistent with one another, and with several independent ways of measuring energy density in cosmology (more about which below). They indicate a non-relativistic matter density of order

$$\rho_{DM0} = 1350 \text{ MeV m}^{-3} \qquad \text{or} \qquad \Omega_{DM0} = 0.26\,. \tag{1.30}$$

The errors in this inference of the size of $\Omega_{DM0}$ are of order 10%. Provided this has the same equation of state, $p \approx 0$, as have the baryons (as is assumed in the ΛCDM model), this leads to a total energy density in non-relativistic matter, $\Omega_{M0} = \Omega_{B0} + \Omega_{DM0}$, which is of order

$$\rho_{M0} = 1600 \text{ MeV m}^{-3} \qquad \text{or} \qquad \Omega_{m0} = 0.30\,. \tag{1.31}$$

**Dark Energy**

Finally, there are also at least two lines of evidence which point to a second form of unknown matter in the universe, independent of the Dark Matter. One line is based on the recent observations that the universal expansion is accelerating, and so requires the universe must now be dominated by a form of matter for which $\rho + 3p < 0$. The second line of argument is based on the observational evidence about the spatial geometry of the universe, which favours the universe being spatially flat, $\kappa = 0$, coming from measurements of the angular fluctuations in the temperature of the CMB. These two lines of evidence are consistent with one another (within sizeable errors) and point to a *Dark Energy* density which is of order

$$\rho_{DE0} = 3600 \text{ MeV m}^{-3} \qquad \text{or} \qquad \Omega_{DE0} = 0.70\,. \tag{1.32}$$

The equation of state for the Dark Energy is not known, apart from the remark that the observations indicate both that at present $\rho_{DE0} \sim 0.7\,\rho_c > 0$ and $w \lesssim -0.7$. If $w$ is constant, it is likely on theoretical grounds that $w = -1$ and the Dark Energy is simply the Lorentz-invariant vacuum energy density. Although it is not yet known whether the vacuum need be Lorentz invariant to the precision required to draw cosmological conclusions of sufficient accuracy, in the ΛCDM model it is assumed that the Dark Energy equation of state is $w = -1$.

### 1.1.5 Earlier epochs

Given the present-day cosmic ingredients of the previous section, it is possible to extrapolate their relative abundances into the past in order to estimate what can be said about earlier cosmic environments. This evolution can be complicated when the various components of the cosmic fluid significantly interact with one another (such as for baryons and photons at redshifts larger than about $z \simeq 1100$, as we shall see), but simplifies immensely if the various components of the cosmic fluid do not exchange stress-energy directly with one another. The ΛCDM model assumes there is no such direct energy exchange between other components and the dark matter and dark energy, and that no exchange exists between the two dark components.

When the component fluids do not directly exchange energy things simplify because eq. (1.12) applies separately to each component individually, dictating the dependence $\rho_i(a)$ and $p_i(a)$ for each of them, as follows:

- **Radiation:** For photons (and relict neutrinos of sufficiently small mass compared with temperature) we have $w = 1/3$ and so $\rho(a)/\rho_0 = (a_0/a)^4$;
- **Non-relativistic Matter:** For both ordinary matter (baryons and electrons) and for the Dark Matter we have $w = 0$ and so $\rho(a)/\rho_0 = (a_0/a)^3$;
- **Vacuum Energy:** Assuming the Dark Energy has the equation of state $w = -1$ we have $\rho(a) = \rho_0$ for all $a$.

This implies the total energy density and pressure have the form

$$\begin{aligned} \rho(a) &= \rho_{DE0} + \rho_{M0}\left(\frac{a_0}{a}\right)^3 + \rho_{R0}\left(\frac{a_0}{a}\right)^4 \\ p(a) &= -\rho_{DE0} + \frac{1}{3}\,\rho_{R0}\left(\frac{a_0}{a}\right)^4 , \end{aligned} \tag{1.33}$$

showing how the relative contribution of each component within the total cosmic fluid changes as it responds differently to the expansion of the universe (see Fig. 1).

As the universe is run backwards to smaller sizes it is clear that the Dark Energy becomes less and less important, while relativistic matter becomes more and more important. Although the Dark Energy is now the dominant contribution to $\rho$ and non-relativistic matter is the next most abundant, when extrapolated backwards they switch roles, so $\rho_M(a) > \rho_{DE}(a)$, relatively recently, at a redshift

$$1 + z = \frac{a_0}{a} > \left(\frac{\Omega_{DE0}}{\Omega_{M0}}\right)^{1/3} \simeq \left(\frac{0.7}{0.3}\right)^{1/3} \simeq 1.3\,. \tag{1.34}$$

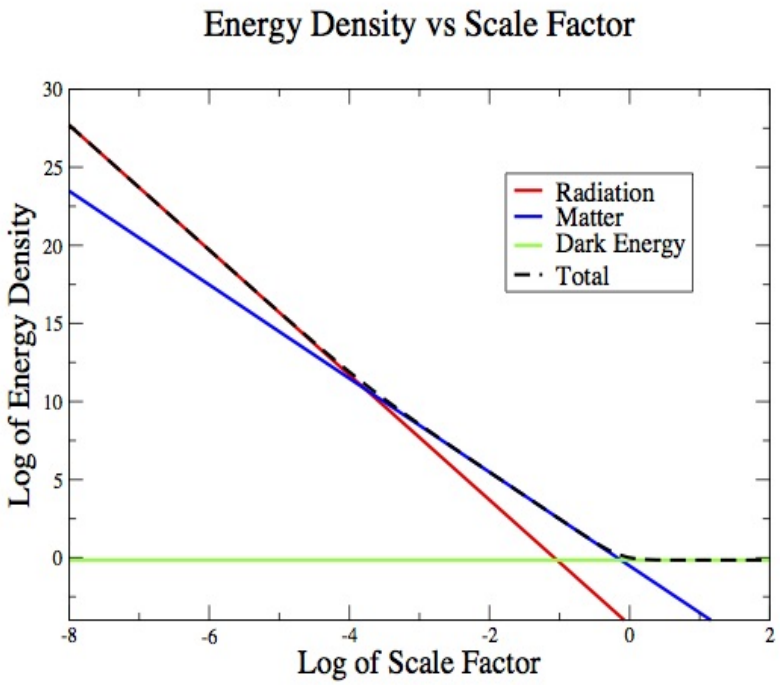


**Figure 1**. The relative abundance (in the energy density) of radiation, nonrelativistic matter and vacuum energy, vs the size of the universe $a/a_0 = (1+z)^{-1}$. The figure assumes negligible direct energy transfer between these fluids, and shows how this implies each type of fluid dominates during particular epochs. The transition from radiation to matter domination (at redshift $z_{\rm eq} \simeq 3600$) plays an important role in the development of structure in the Universe.

In the absence of Dark Matter the energy density in baryons alone would become larger than the Dark Energy density at a slightly earlier epoch

$$1 + z > \left(\frac{\Omega_{DE0}}{\Omega_{B0}}\right)^{1/3} \simeq \left(\frac{0.7}{0.04}\right)^{1/3} \simeq 2.6\,. \tag{1.35}$$

For times earlier than this the dominant component of the energy density is due to non-relativistic matter, and this remains true back until the epoch when the energy density in radiation became comparable with that in non-relativistic matter. Since $\rho_R \propto a^{-4}$ and $\rho_M \propto a^{-3}$ radiation-matter equality occurs when $z = z_{\rm eq}$ with

$$1 + z_{\rm eq} = \frac{\Omega_{M0}}{\Omega_{R0}} \simeq \frac{0.3}{8.4 \times 10^{-5}} \simeq 3600\,. \tag{1.36}$$

This crossover would have occurred much later in the absence of Dark Matter, since the radiation energy density equals the energy density in baryons when

$$1 + z = \frac{\Omega_{B0}}{\Omega_{R0}} \simeq \frac{0.04}{8.4 \times 10^{-5}} \simeq 480\,. \tag{1.37}$$

Knowing how $\rho$ depends on $a$ immediately gives, with the Friedmann equation, $H$ as a function of $a$

$$H(a) = H_0 \Big[\Omega_{DE0} + \Omega_{\kappa 0} \left(\frac{a_0}{a}\right)^2 + \Omega_{M0} \left(\frac{a_0}{a}\right)^3 + \Omega_{R0} \left(\frac{a_0}{a}\right)^4\Big]^{1/2}, \tag{1.38}$$

where we define (as before) $\Omega_f = \rho_f/\rho_c$ for $f =$ radiation$(R)$, matter $(M)$, vacuum $(DE)$ with the critical density defined by $\rho_c := 3H^2 M_p^2$ and the subscript '0' denoting the present epoch.

Eq. (1.38) also defines the curvature contribution to $H$ as

$$\Omega_\kappa := -\frac{\kappa}{(Ha)^2}\,, \tag{1.39}$$

which observations of the CMB (that tell us the Universe is consistent with being spatially flat) tell us is at most of order 10% because the best present-day information indicates that $\Omega_0 = \Omega_{DE0} + \Omega_{m0} + \Omega_{r0} = 1$, which is consistent with $\kappa = 0$. Because $\Omega_\kappa \propto (a_0/a)^2$ it falls more slowly with increasing $a$ than does either matter or radiation. Consequently, given its relatively small size today, $\Omega_\kappa$ contributes negligibly in the remote path and it is a good approximation to take $\kappa = 0$ when discussing the very early Universe.

In principle (1.38) can be inserted into the Friedmann equation and integrated to obtain $a(t)$. Although in general this dependence must be obtained numerically, many of its features follow on simple analytic grounds because for most epochs there is only a single component of the cosmic fluid which is dominating the total energy density. We expect, then, that for redshifts larger than several thousand $a(t) \propto t^{1/2}$ should be a good approximation, as appropriate for the expansion in a universe which is filled purely by radiation. Once $a/a_0$ rises to above 1/3600 there should be a brief transition to the time dependence which describes the universal expansion in a universe dominated by non-relativistic matter and so $a \propto t^{2/3}$. This should apply right up to the very recent past, when $a/a_0$ is around 0.8, after which there is a transition to vacuum-energy domination, during which the universal expansion accelerates to become exponential with $t$. In all likelihood we are at present still living in the transition period from matter to vacuum-energy domination.

### 1.1.6 Thermal evolution

The Hot Big Bang theory of cosmology starts with the idea that the Universe was once small and hot enough that it contained just a soup of elementary particles, in order to see if this leads to a later universe that we recognize in cosmological observations. This picture turns out to describe well many of the features we see around us, which are otherwise harder to understand.

This type of hot fluid cools as the Universe expands, leading to several types of characteristic events whose late-time signatures provide evidence for the validity of the Hot Big Bang picture. The first type of characteristic event is the departure from equilibrium that every species of particle always experiences eventually once its particle density becomes too low for particles to find one another frequently enough to maintain equilibrium.

The second type of characteristic event is the formation of bound states. At finite temperature the net abundance of bound states (like atoms or nuclei, say) is fixed by detailed balance: the competition between reactions (like $e^- p \to H\gamma$) that form the bound states (in this case Hydrogen) and the inverse reactions (like $H\gamma \to e^- p$) that dissociate them. Once the temperature falls below the binding energy of a bound state the typical collision energy falls below the threshold required for dissociation and so the abundance of the bound state grows until the constituents eventually become sufficiently rare that the formation reactions also effectively turn off the production processes. Once this happens the bound-state abundance freezes and for the purposes of later cosmology these bound states can be regarded as being part of the inventory of 'elementary' particles during later epochs.

There is concrete evidence that the formation of bound states took place at least twice in the early Universe. The earliest case happened during the epoch of primordial nucleosynthesis, at redshift $z \simeq 10^{10}$, when temperatures were in the MeV regime and protons and neutrons got cooked into light nuclei. The evidence that this occurred comes from the agreement between the primordial abundances of light nuclear isotopes with the results of precision calculations of their formation rates. Because the total formation rate is proportional to the density of protons and neutrons at this time, the successful agreement between theory and observations also tells us the total density of baryons throughout the Universe at this time.

The second important epoch for forming bound states occurred at the epoch of 'recombination', at redshifts around $z \simeq 1100$, when electrons and nuclei combined to form electrically neutral atoms (like H or He). The evidence for this epoch comes

from the existence and properties of the CMB: the conversion of charged electrons and protons into neutral atoms made the cosmic fluid become transparent to light, as the photons present at that time decoupled from the electron-baryon fluid. These photons continue to rattle around the Universe after this epoch and have been observed. Their distribution has a beautiful thermal form as a function of the present-day photon angular frequency, $\omega_0$, as shown in Fig. 2. The temperature of this distribution has been measured as a function of direction in the sky, $T_\gamma(\theta, \phi)$, and it is the angular average of this measured temperature,

$$T_{\gamma 0} = \langle T_\gamma \rangle = \frac{1}{4\pi} \int T_\gamma(\theta, \phi) \, \sin\theta \, \mathrm{d}\theta \, \mathrm{d}\phi = 2.2725 \text{ K} \,, \tag{1.40}$$

which we use above as the present temperature of the relic photons.

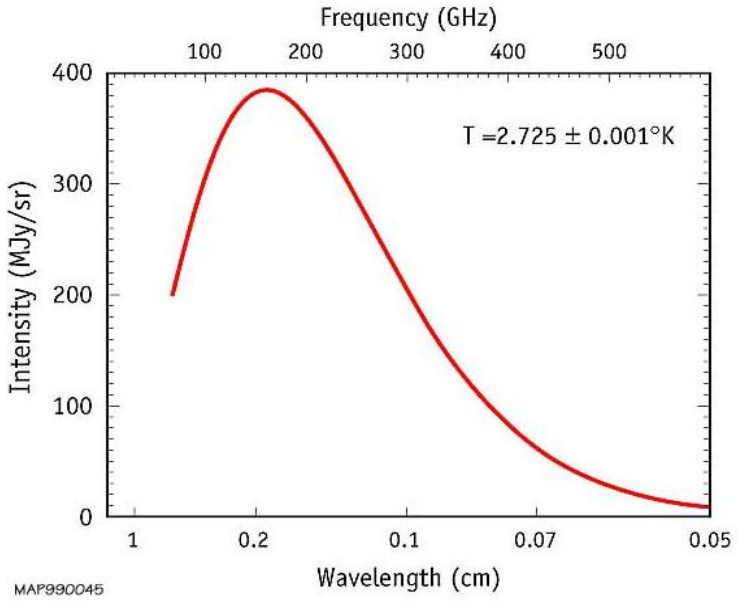


**Figure 2**. The FIRAS measurement of the thermal distribution of the CMB photons. The experimental points lie on the theoretical curve, with errors which are smaller than the width of the curve.

The starting point for making such a thermal description precise is a summary of the various types of particles that are believed to be 'elementary' at the temperatures of interest. The highest temperature for which there is direct observational evidence the universe attained in the past is $T \sim 10^{10}$ K, which corresponds to thermal energies of order 1 MeV. The elementary particles which might be expected to be found within a soup having this temperature are the following.

- **Photons ($\gamma$):** are bosons and have no electric charge or mass, and can be singly emitted and absorbed by any electrically-charged particles.

- **Electrons and Positrons ($e^{\pm}$):** are fermions with charge $\pm e$ and masses equal numerically to $m_e = 0.511$ MeV. Because the positron, $e^+$, is the antiparticle for the electron, $e^-$, (and vice versa), these particles can completely annihilate into photons through the reaction

$$e^+ + e^- \leftrightarrow 2\gamma\,, \tag{1.41}$$

  and do so once the temperature falls below the electron mass.

- **Protons ($p$):** are fermions with charge $+e$ and mass $m_p = 938$ MeV. Unlike all of the other particles described here, the proton and neutron can take part in the strong interactions, for example experiencing reactions like

$$p + n \leftrightarrow D + \gamma\,, \tag{1.42}$$

  in which a proton and neutron combine to produce a deuterium nucleus. The photon which appears in this expression simply carries off any excess energy which is released by the reaction.

- **Neutrons ($n$):** are electrically neutral fermions with mass $m_n = 940$ MeV. Like protons, neutrons participate in the strong interactions. Isolated neutrons are unstable, and left to themselves decay through the weak interactions into a proton, an electron and an electron-antineutrino:

$$n \rightarrow p + e^- + \overline{\nu}_e\,. \tag{1.43}$$

- **Neutrinos and Anti-neutrinos ($\nu_e$, $\overline{\nu}_e$, $\nu_\mu$, $\overline{\nu}_\mu$, $\nu_\tau$, $\overline{\nu}_\tau$):** are fermions that are electrically neutral, and have been found to have nonzero masses whose precise values are not known, but which are known to be smaller than 1 eV.

- **Gravitons ($G$):** are electrically neutral bosons that mediate the gravitational force in the same way that photons do for the electromagnetic force. Gravitons only interact with other particles with gravitational strength, which is much weaker that the strength of the other interactions. As a result they will turn out never to be in thermal equilibrium for any of the temperatures to which we have observational access in cosmology.

To these must be added whatever makes up the Dark Matter, provided temperatures and interactions are such that the Dark Matter can be regarded to be in thermal equilibrium.

How would the temperature of a bath of these particles evolve on thermodynamic grounds as the universe expands? The first step asks how the temperature is related to $a$ (and so also $t$), in order to quantify the rate with which a hot bath cools due to the universal expansion.

**Relativistic Particles**

The energy density and pressure for a gas of relativistic particles (like photons) when in thermal equilibrium at temperature $T_R$ are given by

$$\rho_R = a_B\, T_R^4 \qquad \text{and} \qquad p_R = \frac{1}{3}\, a_B\, T_R^4\,, \tag{1.44}$$

where $a_B$ is $g/2$ times the Stefan-Boltzmann constant and $g$ counts the number of internal (spin) states of the particles of interest (and so $g = 2$ for a gas of photons). Combining this with energy conservation, which says $\rho_R \propto (a_0/a)^4$, shows that the product $aT$ is constant, and so

$$T_R(a) = T_{R0}\left(\frac{a_0}{a}\right) = T_{R0}(1+z)\,. \tag{1.45}$$

This is equivalent to the statement that the expansion is adiabatic, since the entropy per unit volume of a relativistic gas is $s_R \propto T_R^3$, and so the total entropy in this gas is

$$S_R \propto s_R\, a^3 \propto (T_R\, a)^3 = \text{constant}\,. \tag{1.46}$$

Although the relation $T \propto a^{-1}$ is derived above assuming thermal equilibrium, it can continue to hold (for relativistic particles) once the particles become insufficiently dense to scatter frequently enough to maintain equilibrium. This is because the thermal distribution functions for relativistic particles are functions of the ratio of particle energy divided by temperature: $\epsilon/T$. Because relativistic particles have energies $\epsilon(\mathbf{p}) = |\mathbf{p}| = |\mathbf{k}|/a$ their energies redshift $\epsilon \propto a^{-1}$ with the universal expansion. This ensures that the distributions remain in the thermal form for all $t$, provided that their temperature is also regarded as falling with $T \propto a^{-1}$ (so that $\epsilon/T$ is time-independent). For this reason it makes sense to continue to regard the CMB photon temperature to be falling with $T_R \propto a^{-1}$ even though photons stopped interacting frequently enough to remain in equilibrium once protons and electrons combined into electrically neutral atoms around redshift $z \simeq 1100$.

### Nonrelativistic Particles

As mentioned earlier, an ideal gas of non-relativistic particles in thermal equilibrium has a pressure and energy density given instead by

$$p_M = n\,T_M \qquad \text{and} \qquad \rho_M = n\,m + \frac{n\,T_M}{\gamma - 1}\,, \tag{1.47}$$

where $n$ is the number density of particles, $m$ is the particle's rest mass and $\gamma = c_p/c_v$ is its ratio of specific heats, with $\gamma = 5/3$ for a gas of monatomic atoms.

In order to repeat the previous arguments using energy conservation to infer how $T_M$ evolves with $a$ we must first determine what $n$ depends on. If the total number of particles is conserved, so

$$\frac{\mathrm{d}}{\mathrm{d}t}\Big[n\,a^3\Big] = 0\,, \tag{1.48}$$

then consistency of $n \propto a^{-3}$ with energy conservation, eq. (1.12), implies $T_M$ should satisfy

$$\frac{\dot{T}_M}{T_M} + 3(\gamma - 1)\,\frac{\dot{a}}{a} = 0\,, \tag{1.49}$$

and so

$$T_M = T_{M0}\left(\frac{a_0}{a}\right)^{3(\gamma-1)} = T_{M0}(1+z)^{3(\gamma-1)}\,. \tag{1.50}$$

For example, for a monatomic gas with $\gamma = 5/3$ this implies $T_M \propto (1+z)^2 \propto a^{-2}$, as also would be expected for an adiabatic expansion given that the entropy density for such a fluid varies with $T_M$ like $s_M \propto (mT_M)^{3/2}$.

When a nonrelativistic species of particle falls out of equilibrium its energy (because it is nonrelativistic) is dominated by its rest-mass: $\epsilon(\mathbf{p}) \simeq m$. Because of this $\epsilon$ does not redshift and so the distribution of particles remains frozen at the fixed temperature, $T_f$, where equilibrium first broke down.

### Multi-component fluids

The previous examples assume negligible energy exchange between these different components, which in particular also precludes them being in thermal equilibrium with one another (allowing their respective temperatures free to evolve independently of one another). But what happens when several components of the fluid *are* in thermal equilibrium with one another? This situation actually happens for $z > 1100$ when non-relativistic protons and neutrons (or nuclei) are in equilibrium with relativistic photons, electrons and neutrinos.

To see how this works, we now repeat the previous arguments for a fluid which consists of both relativistic and non-relativistic components, coexisting in mutual thermal

equilibrium at a common temperature, $T$. In this case the energy density and pressure are given by

$$p = n\,T + \frac{1}{3}\,a_B\,T^4 \qquad \text{and} \qquad \rho = n\,m + \frac{n\,T}{\gamma - 1} + a_B\,T^4\,. \tag{1.51}$$

Inserting this into the energy conservation equation, as above, leads to the result

$$\frac{\dot{T}}{T} + \left[\frac{1+\sigma}{\sigma + \frac{1}{3}\,(\gamma-1)^{-1}}\right]\,\frac{\dot{a}}{a} = 0\,, \tag{1.52}$$

where

$$\sigma \equiv \frac{4a_B\,T^3}{3n} = 74.0\,\left[\frac{(T/\mathrm{deg})^3}{n/\mathrm{cm}^{-3}}\right]\,, \tag{1.53}$$

is the relativistic entropy per non-relativistic gas particle. For example, if the relativistic gas consists of photons, then the number of photons per unit volume is $n_\gamma = [30\,\zeta(3)/\pi^4]a_B T^3 = 3.7\,a_B T^3$, and so $\sigma = 0.37(n_\gamma/n)$.

Eq. (1.52) shows how $T$ varies with $a$, and reduces to the pure radiation result, $T\,a$ = constant, when $\sigma \gg 1$ and to the non-relativistic matter result, $T\,a^{3(\gamma-1)}$ = constant, when $\sigma \ll 1$. In general, however, this equation has more complicated solutions because $\sigma$ need not be a constant. Given that particle conservation implies $n \propto a^{-3}$, we see that the time-dependence of $\sigma$ is given by $\sigma \propto (T\,a)^3$.

We are led to the following limiting behaviour. If, initially, $\sigma = \sigma_0 \gg 1$ then at early times $T \propto a^{-1}$ and so $\sigma$ remains approximately constant (and large). For such a gas the common temperature of the relativistic and non-relativistic fluids continues to fall like $T \propto a^{-1}$. In this case the high-entropy relativistic fluid controls the temperature evolution and drags the non-relativistic temperature along with it. Interestingly, it can do so even if $\rho_M \approx n\,m$ is larger than $\rho_R = a_B\,T^4$, as can easily happen when $m \gg T$. In practice this happens until the two fluid components fall out of equilibrium with one another, after which their two temperatures continue to evolve separately according to the expressions given previously.

On the other hand if $\sigma = \sigma_0 \ll 1$ initially, then $T \propto a^{-3(\gamma-1)}$ and so $\sigma \propto a^{3(4-3\gamma)}$. This falls as $a$ increases provided $\gamma > 4/3$, and grows otherwise. For instance, the particularly interesting case $\gamma = 5/3$ implies $T \propto a^{-2}$ and so $\sigma \propto a^{-3}$. We see that if $\gamma > 4/3$, then an initially small $\sigma$ gets even smaller still as the universe expands, implying the temperature of both radiation and matter continues to fall like $T \propto a^{-3(\gamma-1)}$. If, however, $1 < \gamma < 4/3$, an initially small $\sigma$ can grow even as the temperature falls, until the fluid eventually crosses over into the relativistic regime for which $T \propto a^{-1}$ and $\sigma$ stops evolving.

## 1.2 An early accelerated epoch

This section now switches from a general description of the ΛCDM model to a discussion about the peculiar initial conditions on which its success seems to rely. This is followed by a summary of the elements of some simple single-field inflationary models, and why their proposal is motivated as explanations of the initial conditions for the later universe.

### 1.2.1 Peculiar initial conditions

The ΛCDM model describes well what we see around us, provided that the Universe is started off with a very specific set of initial conditions. There are several properties of these initial conditions that seem peculiar, as is now summarized.

#### Flatness problem

The first problem concerns the observed spatial flatness of the present-day universe, which is suggested by observations of the temperature fluctuations in the CMB, which indicate that the quantity $\kappa/a^2$ of the Friedmann equation, eq. (1.11), is at present consistent with zero. What is odd about this condition is that this curvature term tends to grow in relative importance as the Universe expands, and finding it to be small now means that it must have been *extremely* small in the remote past.

More quantitatively, it is useful to divide the Friedmann equation by $H^2(t)$ to give

$$1 + \frac{\kappa}{(aH)^2} = \frac{8\pi G\rho}{3H^2} =: \Omega(a)\,, \tag{1.54}$$

where (as before) the final equality defines $\Omega(a)$. The problem arises because the product $aH$ decreases with time during both matter and radiation domination. For instance, observations indicate that at present $\Omega = \Omega_0$ is unity to within about 10%, and since during the matter-dominated era the product $(aH)^2 \propto a^{-1}$ it follows that at the epoch $z_{\rm eq} \simeq 3600$ of radiation-matter equality we must have had

$$\Omega(z_{\rm eq}) - 1 = \Big(\Omega_0 - 1\Big)\left(\frac{a}{a_0}\right) = \frac{\Omega_0 - 1}{1 + z_{\rm eq}} \simeq \frac{0.1}{3600} \simeq 2.8 \times 10^{-5}\,. \tag{1.55}$$

So $\Omega - 1$ had to be smaller than a few tens of a millionth at the time of radiation-matter equality in order to be of order 10% now.

And it only gets worse the further back one goes, provided the extrapolation back occurs within a radiation- or matter-dominated era (as seems to be true at least as far back as the epoch of nucleosynthesis). Since during radiation-domination we have

$(aH)^2 \propto a^{-2}$ and the redshift of nucleosynthesis is $z_{\rm BBN} \sim 10^{10}$ it follows that at this epoch one must require

$$\Omega(z_{\rm BBN}) - 1 = \Big[\Omega(z_{\rm eq}) - 1\Big] \left(\frac{1+z_{\rm eq}}{1+z_{\rm BBN}}\right)^2 = \frac{0.1}{3600}\left(\frac{3600}{10^{10}}\right)^2 \approx 3.6 \times 10^{-18}\,, \quad (1.56)$$

requiring $\Omega$ to be unity with an accuracy of roughly a part in $10^{18}$. The discomfort of having the success of a theory hinge so sensitively on the precise value of an initial condition in this way is known as the Big Bang's *Flatness Problem*.

**Horizon problem**

Perhaps a more serious question asks why the initial universe can be so very homogeneous. In particular, the temperature fluctuations of the CMB only arise at the level of 1 part in $10^5$, and the question is how this temperature can be so incredibly uniform across the sky.

Why is this regarded as a problem? It is not uncommon for materials on earth to have a uniform temperature, and this is usually understood as a consequence of thermal equilibrium because an initially inhomogeneous temperature distribution equilibrates by having heat flow between the hot and cold areas, until everything is eventually all at the same temperature.

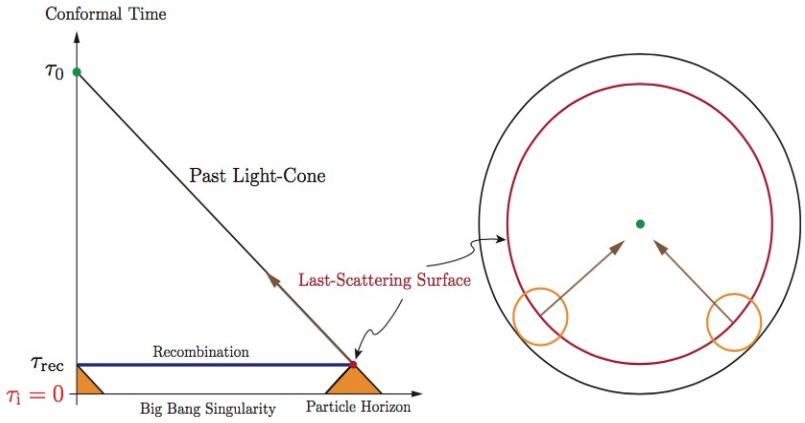


**Figure 3**. A conformal diagram illustrating how there is inadequate time in a radiation-dominated universe for there to be a causal explanation for the correlation of temperature at different points of the sky in the CMB. (Figure taken from [6].)

The same argument is harder to make in cosmology because in the Hot Big Bang model the Universe generically expands so quickly that there has not been enough

time for light to travel across the entire sky to bring everyone the news as to what the common temperature is supposed to be. This is easiest to see using conformal coordinates, as in (1.8), since in these coordinates it is simple to identify which regions can be connected by light signals. In particular, radially directed light rays travel along lines $\mathrm{d}\ell = \pm \mathrm{d}\tau$, which can be drawn as straight lines of slope $\pm 1$ in the $\tau - \ell$ plane, as in Figure 3. The problem is that $a(\tau)$ reaches zero in a finite conformal time (which we can conventionally choose to happen at $\tau = 0$), since $a(\tau) \propto \tau$ during radiation domination and $a(\tau) \propto \tau^2$ during matter domination. Redshift $z_{\rm rec} \simeq 1100$ (the epoch of recombination, at which the CMB photons last sampled the temperature of the Hydrogen gas with which they interact) is simply too early for different directions in the sky to have been causally connected in the entire history of the Universe up to that point.

To pin this down quantitatively, let us assume that the Universe is radiation-dominated for all points earlier than the epoch of radiation-matter equality, $t_{\rm eq}$, so the complete evolution of $a(t)$ until recombination is

$$a(t) \simeq \begin{cases} a_{\rm eq}(t/t_{\rm eq})^{1/2} & \text{for } 0 < t < t_{\rm eq} \\ a_{\rm eq}(t/t_{\rm eq})^{2/3} & \text{for } t_{\rm eq} < t < t_{\rm rec} \,. \end{cases} \tag{1.57}$$

(The real evolution does not have a discontinuous derivative at $t = t_{\rm eq}$, but this inaccuracy is not important for the argument that follows.) The maximum proper distance, measured at the time of recombination, that a light signal could have travelled by the time of recombination, $t_{\rm rec}$, then is

$$\begin{aligned} D_{\rm rec} &= a_{\rm rec} \left[ \int_0^{t_{\rm eq}} \frac{\mathrm{d}\hat{t}}{a(\hat{t})} + \int_{t_{\rm eq}}^{t_{\rm rec}} \frac{\mathrm{d}\hat{t}}{a(\hat{t})} \right] = \frac{a_{\rm rec} t_{\rm eq}}{a_{\rm eq}} \left[ 3 \left( \frac{t_{\rm rec}}{t_{\rm eq}} \right)^{1/3} - 1 \right] \\ &= \frac{2}{H_{\rm eq}^+} \left( \frac{a_{\rm rec}}{a_{\rm eq}} \right)^{3/2} \left[ 1 - \frac{1}{3} \left( \frac{a_{\rm eq}}{a_{\rm rec}} \right)^{1/2} \right] \simeq \frac{1.6}{H_{\rm rec}} \,, \end{aligned} \tag{1.58}$$

where $H_{\rm eq}^+ = 2/(3t_{\rm eq})$ denotes the limit of the Hubble scale as $t \to t_{\rm eq}$ on the matter-dominated side. The approximate equality in this expression uses $H \propto a^{-3/2}$ during matter domination as well as using the redshifts $z_{\rm rec} \simeq 1100$ and $z_{\rm eq} \simeq 3600$ (as would be true in the $\Lambda$CDM model) to obtain $a_{\rm eq}/a_{\rm rec} \simeq 1100/3600 \simeq 0.31$.

To evaluate this numerically we use the present-day value for the Hubble constant, $H_0 \simeq 70$ km/sec/Mpc — or (keeping in mind our units for which $c = 1$), $H_0^{-1} \simeq 13$ Gyr $\simeq 4$ Gpc. This then gives $H_{\rm rec}^{-1} \simeq H_0^{-1}(a_{\rm rec}/a_0)^{3/2} \simeq 3 \times 10^{-5} H_0^{-1} \simeq 0.1$ Mpc, if we use $a_0/a_{\rm rec} = 1 + z_{\rm rec} \simeq 1100$, and so $D_{\rm rec} \simeq 0.2$ Mpc.

Now CMB photons arriving to us from the surface of last scattering left this surface at a distance from us that is now of order

$$R_0 = a_0 \int_{t_{\rm rec}}^{t_0} \frac{{\rm d}\hat{t}}{a(\hat{t})} = 3t_0 - 3t_0^{2/3} t_{\rm rec}^{1/3} = \frac{2}{H_0}\left[1 - \left(\frac{a_{\rm rec}}{a_0}\right)^{1/2}\right], \qquad (1.59)$$

again using $a \propto t^{2/3}$ and $H \propto a^{-3/2}$, and so $R_0 \simeq 2/H_0 \simeq 8$ Gpc. So the angle subtended by $D_{\rm rec}$ placed at this distance away (in a spatially-flat geometry) is really $\theta \simeq D_{\rm rec}/R_{\rm rec}$ where $R_{\rm rec} = (a_{\rm rec}/a_0)R_0 \simeq 7$ Mpc is its distance *at the time of last scattering*, leading to $\theta \simeq 0.2/7 \simeq 1^o$. Any two directions in the sky separated by more than this angle (about twice the angular size of the Moon, seen from Earth) are so far apart that light had not yet had time to reach one from the other since the universe's beginning.

How can all the directions we see have known they were all to equilibrate to the same temperature? It is very much as if we were to find a very uniform temperature distribution, *immediately* after the explosion of a very powerful bomb.

**Defect problem**

Historically, a third problem — called the 'Defect' (or 'Monopole') Problem is also used to motivate changing the extrapolation of radiation domination into the remote past. A defect problem arises if the physics of the much higher energy scales relevant to the extrapolation involves the production of topological defects, like domain walls, cosmic strings or magnetic monopoles. Such defects are often found in Grand Unified theories; models proposed to unify the strong and electroweak interactions as energies of order $10^{15}$ GeV.

These kinds of topological defects can be fatal to the success of late-time cosmology, depending on how many of them survive down to the present epoch. For instance if the defects are monopoles, then they typically are extremely massive and so behave like non-relativistic matter. This can cause problems if they are too abundant because they can preclude the existence of a radiation dominated epoch, because their energy density falls more slowly than does radiation as the universe expands.

Defects are typically produced with an abundance of one per Hubble volume, $n_d(a_f) \sim H_f^3$, where $H_f = H(a_f)$ is the Hubble scale at their epoch of formation, at which time $a = a_f$. Once produced, their number is conserved, so their density at later times falls like $n_d(a) = H_f^3 (a_f/a)^3$. Consequently, at present the number surviving within a Hubble volume is $n_d(a_0) H_0^{-3} = (H_f\, a_f / H_0\, a_0)^3$.

Because the product $aH$ is a falling function of time, the present-day abundance of defects can easily be so numerous that they come to dominate the universe well before

the nucleosynthesis epoch.[7] This could cause the universe to expand (and so cool) too quickly as nuclei were forming, and so give the wrong abundances of light nuclei. Even if not sufficiently abundant during nucleosynthesis, the energy density in relict defects can be inconsistent with measures of the current energy density.

This is clearly more of a hypothetical problem than are the other two, since whether there is a problem depends on whether the particular theory for the high-energy physics of the very early universe produces these types of defects or not. It can be fairly pressing in Grand Unified models since in these models the production of magnetic monopoles can be fairly generic.

### 1.2.2 Acceleration to the rescue

The key observation when trying to understand the above initial conditions is that they only seem unreasonable because they are based on extrapolating into the past assuming the Universe to be radiation (or matter) dominated (as would naturally be true if the ΛCDM model were the whole story). This section argues that these initial conditions can seem more reasonable if a different type of extrapolation is used; in particular if there were an earlier epoch during which the Universal expansion were to accelerate: $\ddot{a} > 0$.

Why should acceleration help? The key point is that what made the above initial conditions a problem was the fact that the product $aH$ is a falling function as $a$ increases, for both matter and radiation domination. But if $\ddot{a} > 0$ then $\dot{a} = aH$ *increases* as $a$ increases, and this can help alleviate the problems.

Why does it matter whether $aH$ increases or decreases? This is perhaps easiest to see for the flatness problem, since this problem relies on the evolution: $\Omega - 1 \propto (aH)^{-2}$. This is a growing function only if $aH$ decreases with time, and so $\Omega - 1$ need not be unusually small in the past if there is a sufficiently long epoch before nucleosynthesis during which $aH$ were to grow.

How long is long enough? To pin this down suppose there were an earlier epoch during which the Universe were to expand in the same way as during Dark Energy domination, $a(t) \propto e^{Ht}$, for constant $H$. Then $aH = a_0 H\, e^{Ht}$ grows exponentially with time and so even if $Ht$ were of order 100 or less it would be possible to explain why $\Omega - 1$ could be as small as $10^{-18}$ or smaller.

[7] Whether they do also depends on their dimension, with magnetic monopoles tending to be more dangerous in this regard than are cosmic strings, say.

Having $aH$ grow also allows a resolution to the horizon problem. One way to see this is to notice that $a(t) \propto e^{Ht}$ implies $\tau \propto e^{-Ht}$ and so

$$a(\tau) = -\frac{1}{H\tau}\,, \tag{1.60}$$

with $0 < a < \infty$ corresponding to the range $-\infty < \tau < 0$. Exponentially accelerated expansion allows $\tau$ to be extrapolated into negative values, and so allows sufficient time for the two causally disconnected regions of the conformal diagram of Figure 3 to have at one point been in causal contact.

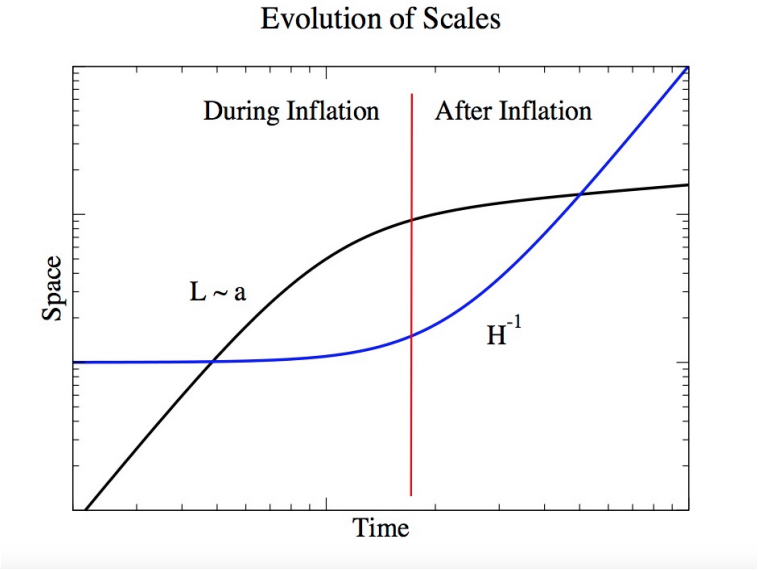


**Figure 4**. A sketch of the relative growth of physical scales, $L(t)$, (in black) and the Hubble length, $H^{-1}$, (in blue) during and after inflation. Horizon exit happens during inflation where the blue and black curves first cross, and this is eventually followed by horizon re-entry where the curves cross again during the later Hot Big Bang era.

Another way to visualize this is to plot physical distance $\lambda(t) \propto a(t)$ and the Hubble radius, $H^{-1}$, against $t$, as in Figure 4. During radiation or matter domination we have $H^{-1} \propto t$ while $a(t) \propto t^p$ with $0 < p < 1$, and so $H^{-1}$ grows more quickly with $t$ than do physical length scales $\lambda(t)$. The causality problem arises because physical quantities tend to freeze when their corresponding length scales satisfy $\lambda(t) > H^{-1}$, thereby precluding physical processes to act over these scales to explain things like the uniform temperature of the CMB. During radiation or matter domination systems of

any given size eventually get caught by the growth of $H^{-1}$ and so 'come inside the Hubble scale' as the Universe expands. Systems iinvolving larger $\lambda(t)$ do so later than those with smaller $\lambda$ and the largest sizes visible have only recently done so and so cannot have been evolving at all over the history of a radiation (or matter) dominated universe.

The freezing of super-Hubble scales can be seen, for example, in the evolution of a massless scalar field in an expanding universe, since the field equation $\Box\phi = 0$ becomes in FRW coordinates

$$\ddot{\phi}_k + 3H\,\dot{\phi}_k + \left(\frac{k}{a}\right)^2 \phi_k = 0\,, \tag{1.61}$$

where we Fourier expand the field $\phi(x) = \int \mathrm{d}^3k\ \phi_k \exp\left[i\mathbf{k}\cdot\mathbf{x}\right]$ using co-moving coordinates, $\mathbf{x}$. For modes satisfying $2\pi/\lambda = p = k/a \ll H$ the field equation implies $\dot{\phi}_k \propto a^{-3}$ and so $\phi_k = C_0 + C_1 \int \mathrm{d}t/a^3$ is the sum of a constant plus a decaying mode.

Things are very different during exponential expansion, however, since $\lambda(t) \propto a(t) \propto e^{Ht}$ grows exponentially with $t$ while $H^{-1}$ remains constant. This means that modes that are initially smaller than the Hubble length get stretched to become larger than the Hubble length, with the transition for a specific mode of length $\lambda(t)$ occurring at the epoch of 'Hubble exit', $t = t_{\rm he}$, defined by $2\pi/\lambda(t_{\rm he}) = p_{\rm he} = k/a(t_{\rm he}) = H$.

It is because the criterion for Hubble exit and entry is $k = aH$ that the growth or shrinkage of $aH$ is relevant to the horizon problem.

How much expansion is required to solve the horizon problem? Choosing a mode $\phi_k$ that is only now crossing the Hubble scale tells us that $k = a_0 H_0$. This same mode would have crossed the horizon during an exponentially expanding epoch when $k = a_{\rm he} H_I$, where $H_I$ is the constant Hubble scale during exponential expansion. So clearly $a_0 H_0 = a_{\rm he} H_I$ where $t_{\rm he}$ is the time of exit for this particular mode. To determine how much exponential expansion is required we solve the following equation for $N_e := \ln(a_{\rm end}/a_{\rm he})$, where $a_{\rm end}$ is the scale factor at the end of the exponentially expanding epoch.

$$1 = \frac{a_{\rm he}H_I}{a_0 H_0} = \left(\frac{a_{\rm he}H_I}{a_{\rm end}H_I}\right)\left(\frac{a_{\rm end}H_I}{a_{\rm eq}H_{\rm eq}}\right)\left(\frac{a_{\rm eq}H_{\rm eq}}{a_0 H_0}\right) = e^{-N_e}\left(\frac{a_{\rm eq}}{a_{\rm end}}\right)\left(\frac{a_0}{a_{\rm eq}}\right)^{1/2}\,, \tag{1.62}$$

which assumes (for the purposes of argument) that the Universe is radiation dominated right from $t_{\rm end}$ until radiation-matter equality, and uses $aH \propto a^{-1}$ during radiation domination and $aH \propto a^{-1/2}$ during matter domination. $N_e = H_I(t_{\rm end} - t_{\rm he})$ is called the number of $e$-foldings of exponential expansion and is proportional to how long exponential expansion lasts

Using, as above, $(a_{\rm eq}H_{\rm eq})/(a_0H_0) = (a_0/a_{\rm eq})^{1/2} \simeq 60$, and $(a_{\rm eq}H_{\rm eq})/(a_{\rm end}H_{\rm end}) = a_{\rm end}/a_{\rm eq} = T_{\rm eq}/T_M$ with $T_{\rm eq} \sim 3$ eV, and assuming the energy density of the exponentially expanding phase is transferred perfectly efficiently to produce a photon temperature $T_M$ then leads to the estimate

$$N_e \sim \ln\left[(3\times 10^{23})\times 60\right] + \ln\left(\frac{T_M}{10^{15}\ {\rm GeV}}\right) \approx 58 + \ln\left(\frac{T_M}{10^{15}\ {\rm GeV}}\right) . \qquad (1.63)$$

Roughly 60 $e$-foldings of exponential expansion can provide a framework for explaining how causal physics might provide the observed correlations that are observed in the CMB over the largest scales, even if the energy densities involved are as high as $10^{15}$ GeV. We shall see below that life is even better than this, because in addition to providing a *framework* in which a causal understanding of correlations could be solved, inflation itself can provide the *mechanism* for explaining these correlations (given an inflationary scale of the right size).

### 1.2.3 Inflation or a bounce?

An early epoch of near-exponential accelerated expansion has come to be known as an 'inflationary' early Universe. Acceleration within this framework speeds up an initially expanding Universe to a higher expansion rate. However, an attentive reader may notice that although acceleration is key to helping with ΛCDM's initial condition issues, there is no *a priori* reason why the acceleration must occur in an initially expanding universe, as opposed (say) to one that is initially contracting. Models in which one tries to solve the problems of ΛCDM by having an initially contracting universe accelerate to become an expanding one are called 'bouncing' cosmologies.

Since it is really the acceleration that is important, bouncing models should in principle be on a similar footing to inflationary ones. In what follows only inflationary models are considered, for the following reasons:

#### Validity of the semiclassical methods

Predictions in essentially all cosmological models are extracted using semiclassical methods: one typically writes down the action for some system and then explores its consequences by solving its classical equations of motion. So a key question for all such models is the identification of the small parameter (or parameters) that suppresses quantum effects and so controls the underlying semiclassical approximation. In the absence of such a control parameter classical predictions need not capture what the system really does. Such a breakdown of the semiclassical approximation really means

that the 'theory error' in the model's predictions could be arbitrarily large, making comparisons to observations essentially meaningless.

A reason sometimes given for not pinning down the size of quantum corrections when doing cosmology is that gravity plays a central role, and we do not yet know the ultimate theory of quantum gravity. Implicit in this argument is the belief that the size of quantum corrections is incalculable without such an ultimate theory, such as due to the well-known divergences in quantum predictions due to the non-renormalizability of General Relativity. But experience with non-renormalizable interactions elsewhere in physics tells us that quantum predictions can sometimes be made, provided one recognizes they involve an implicit low-energy/long-distance expansion relative to the underlying physical scale set by the dimensionful non-renormalizable couplings. Because of this the semiclassical expansion parameter in such theories is usually the ratio between this underlying short-distance scale and the distances of interest in cosmology (which, happily enough, aims at understanding the largest distances on offer). Effective field theories provide the general tools for quantifying these low-energy expansions, and this is why EFT methods are so important for any cosmological studies.

As is argued in more detail in §3, the semiclassical expansion in cosmology is controlled by small quantities like $(\lambda M_p)^{-2}$ where $\lambda$ is the smallest length scale associated with the geometry of interest. In practice it is often $\lambda \sim H^{-1}$ that provides the relevant scale in cosmology, particularly when all geometrical dimensions are similar in size. So a rule of thumb generically asks the ratio $H^2/M_p^2$ to be chosen to be small:

$$\frac{H^2}{M_p^2} \propto \frac{\rho}{M_p^4} \ll 1\,, \tag{1.64}$$

as a necessary condition[8] for quantum cosmological effects to be suppressed.

For inflationary models $H$ is usually at its largest during the inflationary epoch, with geometrical length scales only increasing thereafter, putting one deeper and deeper into the semiclassical domain. It is a big plus for these models that they can account for observations while wholly remaining within the regime set by (1.64), and this is one of the main reasons why they receive so much attention.

[8]The semiclassical criterion can be stronger than this, though this can often only be quantified within the context of a specific proposal for what quantum gravity is at the shortest scales. For instance, if it is string theory that takes over at the shortest scales then treatment of cosmology using a field theory – rather than fully within string theory – requires (1.64) be replaced by the stronger condition $H^2/M_s^2 \ll 1$, where $M_s \ll M_p$ is the string scale, set for example by the masses of the lightest string excited states.

For bouncing cosmologies the situation can be more complicated. The smallest geometrical scale $\lambda$ usually occurs during the epoch near the bounce, even though $H^{-1}$ itself usually tends to infinity there. In models where $\lambda$ becomes comparable to $M_p$ (or whatever other scale – such as the string scale, $M_s \ll M_p$ – that governs short-distance gravity), quantum effects during the bounce need not be negligible and the burden on proponents is to justify why semiclassical predictions actually capture what happens during the bounce.

**Difficulty of achieving a semiclassically large bounce**

Another issue arises even if the scale $\lambda$ during a bounce does remain much larger than the more microscopic scales of gravity. In this regime the bounce can be understood purely within the low-energy effective theory describing the cosmology, for which General Relativity should be the leading approximation. But (when $\kappa = 0$) the Friedmann equation for FRW geometries in General Relativity states that $H^2 = \rho/3M_p^2$, and so $\rho$ must pass through zero at the instant where the contracting geometry transitions to expansion (since $H = \dot{a}/a$ vanishes at this point). Furthermore, using (1.11) and (1.13), it must also be true that

$$\dot{H} = \frac{\ddot{a}}{a} - H^2 = -\frac{1}{2M_p^2}(\rho + p) > 0\,, \tag{1.65}$$

at this point in order for $H$ to change sign there, which means the dominant contributions to the cosmic fluid must satisfy $\rho + p < 0$ during the bounce.[9]

Although there are no definitive no-go theorems, it has proven remarkably difficult to find a convincing physical system that both satisfies the condition $\rho + p < 0$ and does not also have other pathologies, such as uncontrolled runaway instabilities. For instance within the class of multiple scalar field models for which the lagrangian density is $\mathcal{L} = \sqrt{-g}\left[\frac{1}{2}\,G_{ij}(\phi)\,\partial_\mu\phi^i\,\partial^\mu\phi^j + V(\phi)\right]$ we have $\rho + p = G_{ij}(\phi)\,\dot{\phi}^i\,\dot{\phi}^j$ and so $\rho + p < 0$ requires the matrix of functions $G_{ij}(\phi)$ to have a negative eigenvalue. But if this is true then there is always a combination of fields for which the kinetic energy is negative (what is called a 'ghost'), and so is unstable towards the development of arbitrarily rapid motion.

**Phenomenological issues**

In addition to the above conceptual issues involving the control of predictions, there are also potential phenomenological issues that bouncing cosmologies must face. Whereas

[9] This is usually phrased as a violation of the 'null-energy' condition, which states that $T_{\mu\nu}n^\mu n^\nu \geq 0$ for all null vectors $n^\mu$.

expanding geometries tend to damp out spatially varying fluctuations – such as when gradient energies involve factors like $(k/a)^2$ that tend to zero as $a(t)$ grows – the opposite typically occurs during a contracting epoch for which $a(t)$ shrinks. This implies that inhomogeneities tend to grow during the pre-bounce contraction, and so a mechanism must be provided for the emergence into the homogeneous and isotropic later universe we see around us in observational cosmology.

It is of course important that bouncing cosmologies be investigated, not least in order to see most fully what might be required to understand the flatness and horizon problems and whether there are alternative observational implications to those of inflation that might be used to marshal evidence about what actually occurred in the very early universe. But within the present state of the art inflationary models have one crucial advantage over bouncing cosmologies: they provide concrete semiclassical control over the key epoch of acceleration on which the success of the model ultimately relies. Because of this inflationary models are likely to remain the main paradigm for studying pre-ΛCDM extrapolations, at least until bouncing cosmologies are developed to allow similar control over how primordial conditions get propagated to the later universe through the bounce.

### 1.2.4 Simple inflationary models

So far so good, but what kind of physics can provide both an early period of accelerated expansion and a mechanism for ending this expansion to allow for the later emergence of the successful Hot Big Bang cosmology?

Obtaining the benefits of an exponential expansion requires two things: $(i)$ some sort of physics that hangs the universe up for a relatively long period with an accelerating equation of state, $p < -\frac{1}{3}\rho < 0$; *and* $(ii)$ some mechanism for ending this epoch to allow the later appearance of the radiation-dominated epoch within which the usual Big Bang cosmology starts. Although a number of models exist that can do this, none yet seems completely compelling. This section describes some of the very simplest such models.

The central requirement is to have some field temporarily dominate the universe with potential energy, and for the vast majority of models this new physics comes from the dynamics of a scalar field, $\varphi(x)$, called the 'inflaton'. This field can be thought of as an order parameter characterizing the dynamics of the vacuum at the very high energies likely to be relevant to inflationary cosmology. Although the field $\varphi$ can in principle depend on both position and time, inflation turns out rapidly to smooth out spatial variations, and so it suffices to study $\varphi = \varphi(t)$.

No way is known to obtain a viable inflationary model simply using the known particles and interactions, but a minimal model [7] does use the usual scalar Higgs field already present in the Standard Model as the inflaton, provided it is assumed to have a nonminimal coupling to gravity of the form $\delta\mathcal{L} = -\xi\,\sqrt{-g}\;(\mathcal{H}^\dagger\mathcal{H})\,R$, where $\mathcal{H}$ is the usual Higgs doublet and $R$ is the Ricci scalar. Here $\xi$ is a new dimensionless coupling, whose value turns out must be of order $10^4$ in order to provide a good description of cosmological observations. Inflation in this case turns out to occur when the Higgs field takes values out at trans-Planckian values, $\mathcal{H}^\dagger\mathcal{H} > M_p^2$, assuming $V$ remains approximately proportional to $(\mathcal{H}^\dagger\mathcal{H})^2$ at such large values.

As argued in [8], although the large values required for both $\xi$ and $\mathcal{H}^\dagger\mathcal{H}$ needn't invalidate the validity of the EFT description, they do push the envelope for the boundaries of its domain of validity. In particular, semiclassical expansion during inflation turns out to require the neglect of powers of $\sqrt{\xi}H/M_p$, which during inflation turns out is to be evaluated with $H \sim M_p/\xi$.

The simplest models instead propose a single new relativistic scalar field, $\varphi$, and designs its dynamics through choices made for its potential energy, $V(\varphi)$. Taking

$$\mathcal{L} = \sqrt{-g}\left[\frac{1}{2}\,\partial_\mu\varphi\,\partial^\mu\varphi + V(\varphi)\right], \tag{1.66}$$

the inflaton field equation becomes $\Box\varphi = V'(\varphi)$, which for homogeneous configurations $\varphi(t)$ reduces to

$$\ddot{\varphi} + 3H\dot{\varphi} + V' = 0\,, \tag{1.67}$$

where $V' = \mathrm{d}V/\mathrm{d}\varphi$.

The Einstein field equations are as before, but with new $\varphi$-dependent contributions to the energy density and pressure: $\rho = \rho_{\rm rad} + \rho_{\rm m} + \rho_\varphi$ and $p = \frac{1}{3}\,\rho_{\rm rad} + p_\varphi$, where

$$\rho_\varphi = \frac{1}{2}\,\dot{\varphi}^2 + V(\varphi) \qquad \text{and} \qquad p_\varphi = \frac{1}{2}\,\dot{\varphi}^2 - V(\varphi)\,. \tag{1.68}$$

The Dark Energy of the present-day epoch is imagined to arise by choosing $V$ so that its minimum satisfies $\rho_{DE} = V(\varphi_{\rm min})$. Inflation is imagined to occur when $\varphi$ evolves slowly through a region where $V(\varphi) \gg V(\varphi_{\rm min})$ is very large, and ends once $\varphi$ rolls down towards its minimum.

With these choices energy conservation for the $\varphi$ field — $\dot{\rho}_\varphi + 3(\dot{a}/a)(\rho_\varphi + p_\varphi) = 0$ follows from the field equation, eq. (1.67). Some couplings must also exist between the $\varphi$ field and ordinary Standard Model particles in order to provide a channel to transfer energy from the inflaton to ordinary particles, and so reheat the universe as required for

the later Hot Big Bang cosmology. But $\varphi$ is not imagined to be in thermal equilibrium with itself or with the other kinds of matter during inflation or at very late times, and this can be self-consistent if the coupling to other matter is sufficiently weak and if the $\varphi$ particles are too heavy to be present once the cosmic fluid cools to the MeV energies and below (for which we have direct observations).

### Slow-Roll Inflation

To achieve an epoch of near-exponential expansion, we seek a solution to the above classical field equations for $\varphi(t)$ in which the Hubble parameter, $H$, is approximately constant. This is ensured if the total energy density is dominated by $\rho_\varphi$, with $\rho_\varphi$ also approximately constant. As we have seen, energy conservation implies the pressure must then satisfy $p_\varphi \approx -\rho_\varphi$. Inspection of eqs. (1.68) shows that both of these conditions are satisfied if the $\varphi$ kinetic energy is negligible compared with its potential energy:

$$\frac{1}{2}\dot{\varphi}^2 \ll V(\varphi)\,, \tag{1.69}$$

since then $p_\varphi \simeq -V(\varphi) \simeq -\rho_\varphi$. So long as $V(\varphi)$ is also much larger than any other energy densities, it would dominate and $H^2 \simeq V/(3M_p^2)$ would then be approximately constant.

What properties must $V(\varphi)$ satisfy in order to allow (1.69) to hold for a sufficiently long time? This requires a long period of time where $\varphi$ moves slowly enough to allow *both* the neglect of $\frac{1}{2}\dot{\varphi}^2$ relative to $V(\varphi)$ in the Friedmann equation, (1.11), *and* the neglect of $\ddot{\varphi}$ in the scalar field equation, (1.67).

The second of these conditions allows eq. (1.67) to be written in the approximate *slow-roll* form,

$$\dot{\varphi} \approx -\left(\frac{V'}{3H}\right)\,. \tag{1.70}$$

Using this in (1.69) then shows $V$ must satisfy $(V')^2/(9H^2V) \ll 1$, leading to the condition that slow-roll inflation requires $\varphi$ must lie in a region for which

$$\epsilon := \frac{1}{2}\left(\frac{M_p V'}{V}\right)^2 \ll 1\,. \tag{1.71}$$

Physically, this condition requires $H$ to be approximately constant over any given Hubble time, inasmuch as $3M_p^2H^2 \simeq V$ implies $6M_p^2H\dot{H} \simeq V'\dot{\varphi} \simeq -(V')^2/3H$ and so

$$-\frac{\dot{H}}{H^2} \simeq \frac{(V')^2}{18H^4M_p^2} \simeq \frac{M_p^2(V')^2}{2V^2} = \epsilon \ll 1\,. \tag{1.72}$$

Self-consistency also demands that if eq. (1.70) is differentiated to compute $\ddot{\varphi}$ it should be much smaller than $3H\dot{\varphi}$. Performing this differentiation and demanding that $\ddot{\varphi}$ remain small (in absolute value) compared with $3H\dot{\varphi}$, then implies $|\eta| \ll 1$ where

$$\eta := \frac{M_p^2\, V''}{V}\,, \tag{1.73}$$

defines the second slow-roll parameter. The slow-roll parameters $\epsilon$ and $\eta$ are important because (as we see below) the key predictions of single-field slow-roll inflation for density fluctuations can be expressed in terms of the three parameters $\epsilon$, $\eta$ and the value, $H_I$, of the Hubble parameter during inflation.

Given an explicit shape for $V(\varphi)$ one can directly predict the amount of inflation that occurs between the end of inflation and the epoch of horizon exit where the scales of interest become larger than the Hubble length. This is done by relating the amount of expansion directly to the distance $\varphi$ traverses in field space between these two epochs. To this end, rewriting eq. (1.70) in terms of $\varphi' \equiv \mathrm{d}\varphi/\mathrm{d}a$, leads to

$$\frac{\mathrm{d}\varphi}{\mathrm{d}a} = \frac{\dot{\varphi}}{\dot{a}} = -\,\frac{V'}{3aH^2} = -\,\frac{M_p^2\, V'}{aV}\,, \tag{1.74}$$

which when integrated between horizon exit, $\varphi_{\rm he}$, and final value, $\varphi_{\rm end}$, gives the amount of expansion during inflation as $a_{\rm end}/a_{\rm he} = e^{N_e}$, with

$$N_e = \int_{a_{\rm he}}^{a_{\rm end}} \frac{\mathrm{d}a}{a} = \int_{\varphi_{\rm end}}^{\varphi_{\rm he}} \mathrm{d}\varphi \left(\frac{V}{M_p^2\, V'}\right) = \frac{1}{M_p}\int_{\varphi_{\rm end}}^{\varphi_{\rm he}} \frac{\mathrm{d}\varphi}{\sqrt{2\epsilon}}\,. \tag{1.75}$$

In these expressions $\varphi_{\rm end}$ can be defined by the point where the slow-roll parameters are no longer small, such as where $\epsilon \simeq \frac{1}{2}$. Then this last equation can be read as defining $\varphi_{\rm end}(N_e)$, as a function of the desired number of $e$-foldings between the the epoch of horizon exit and the end of inflation, since this is this quantity constrained to be large by the horizon and flatness problems.

Notice also that if $\epsilon$ were approximately constant during inflation, then eq. (1.75) implies that $N_e \approx (\varphi_{\rm he} - \varphi_{\rm end})/(\sqrt{2\epsilon}\, M_p)$. In such a case $\varphi$ must traverse a range of order $N_e M_p \sqrt{2\epsilon}$ between $\varphi_{\rm he}$ and $\varphi_{\rm end}$. This is larger than order $M_p$ provided only that $1 \gg \epsilon \gtrsim 1/N_e^2$, indicating why it is often large fields that are of interest for inflation.

It is worth working through what these formulae mean in a few concrete choices for the shape of the scalar potential.

### Example I: Quadratic model

The simplest example of an inflating potential chooses $\varphi$ to be a free massive field, for which

$$V = \frac{1}{2}\, m^2\, \varphi^2\,, \tag{1.76}$$

and so $V' = m^2\,\varphi$ and $V'' = m^2$, leading to slow-roll parameters of the form

$$\epsilon = \frac{1}{2}\left(\frac{2M_p}{\varphi}\right)^2 \qquad \text{and} \qquad \eta = \frac{2M_p^2}{\varphi^2}\,, \tag{1.77}$$

and so $\epsilon = \eta$ in this particular case, and slow roll requires $\varphi \gg M_p$. The scale for inflation in this field range is $V = \frac{1}{2}\,m^2\,\varphi^2$ and so $H_I^2 \simeq m^2\,\varphi^2/(6\,M_p^2)$. We can ensure $H_I^2/M_p^2 \ll 1$ even if $\varphi \gg M_p$ by choosing $m/M_p$ sufficiently small. Observations will turn out to require $\epsilon \sim \eta \sim 0.01$ and so the regime of interest is $\varphi_{\rm he} \sim 10M_p$, and so small $H_I/M_p$ requires $m/M_p \ll 0.1$.

In this regime $\varphi$ (and so also $V$ and $H$) remains approximately constant despite there being no stationary point for $V$ at large $\varphi$ because Hubble friction keeps $\varphi$ from sliding down the potential very quickly. Since $\varphi$ evolves towards smaller values, eventually slow roll ends once $\eta$ and $\epsilon$ become $O(1)$. Choosing $\varphi_{\rm end}$ by the condition $\epsilon(\varphi_{\rm end}) = \eta(\varphi_{\rm end}) = \frac{1}{2}$ implies $\varphi_{\rm end} = 2M_p$. The number of $e$-foldings between horizon exit and $\varphi_{\rm end} = 2M_p$ is then given by eq. (1.75), which in this instance becomes

$$N_e = \int_{2M_p}^{\varphi_{\rm he}} {\rm d}\varphi \left(\frac{\varphi}{2M_p^2}\right) = \left(\frac{\varphi_{\rm he}}{2M_p}\right)^2 - 1\,, \tag{1.78}$$

and so obtaining $N_e \sim 63$ $e$-foldings (say) requires choosing $\varphi_{\rm he} \sim 16\,M_p$. In particular $\epsilon_{\rm he} := \epsilon(\varphi_{\rm he})$ and $\eta_{\rm he} := \eta(\varphi_{\rm he})$ can be expressed directly in terms of $N_e$, leading to

$$\epsilon_{\rm he} = \eta_{\rm he} = \frac{1}{2(N_e+1)}\,, \tag{1.79}$$

which are both of order $10^{-2}$ for $N_e \simeq 60$.

**Example II: pseudo-Goldstone axion**

From the point of view of particle physics it is more natural to suppose the inflaton is a pseudo-Goldstone boson because then its mass is protected by an approximate shift symmetry. The need for this kind of protection arises because the condition $|\eta| \ll 1$ implies the inflaton mass must be very small compared with the other scales during inflation, because $m^2 \sim |V''| \sim |\eta\,V/M_p^2| \ll H^2$.

If the approximate shift symmetry arises as a phase rotation for some field, and if the symmetry under continuous shifts is broken down to discrete shifts, then it is natural to suppose the scalar potential should be trigonometric:

$$V = V_0 + \Lambda^4\left[1 - \cos\left(\frac{\varphi}{f}\right)\right] = V_0 + 2\Lambda^4 \sin^2\left(\frac{\varphi}{2f}\right)\,, \tag{1.80}$$

for some scales $V_0$, $\Lambda$ and $f$. Here $V_0$ is chosen to agree with $\rho_{DE}$, while the scales $\Lambda$ and $f$ are dictated by the requirements of inflation. Because $\rho_{DE}$ is so small the parameter $V_0$ is dropped in what follows.

With this choice $V' = (\Lambda^4/f)\sin(\varphi/f)$ and $V'' = (\Lambda^4/f^2)\cos(\varphi/f)$, leading to slow-roll parameters of the form

$$\epsilon = \frac{M_p^2}{8f^2}\,\cot^2\left(\frac{\varphi}{2f}\right) \qquad \text{and} \qquad \eta = \frac{M_p^2}{2f^2}\left[\cot^2\left(\frac{\varphi}{2f}\right) - 1\right], \tag{1.81}$$

and so $\eta = 4\epsilon - (M_p^2/2f^2)$. Notice that in the limit $\varphi \ll f$ these go over to the $m^2\varphi^2$ case examined above, with $m \simeq \Lambda^2/f$.

Slow roll in this model typically requires $f \gg M_p$. This can be seen directly from (1.84) for generic $\varphi \simeq f$, but also follows when $\varphi \ll f$ because in this case the potential is close to quadratic and slow roll requires $M_p \ll \varphi \ll f$. The scale for inflation is $V \simeq \Lambda^4$ and so $H_I \sim \Lambda^2/M_p$, and so $H_I^2/M_p^2 \ll 1$ provided we take $\Lambda \ll M_p$. Obtaining $\epsilon \sim \eta \sim 0.01$ can be arranged by choosing $f \sim 10M_p$.

The number of $e$-foldings between horizon exit and $\varphi_{\rm end}$ is again given by eq. (1.75), so

$$N_e = \frac{2f}{M_p^2}\int_{\varphi_{\rm end}}^{\varphi_{\rm he}} \mathrm{d}\varphi\ \tan\left(\frac{\varphi}{2f}\right) = \left(\frac{2f}{M_p}\right)^2 \ln\left|\frac{\sin(\varphi_{\rm he}/2f)}{\sin(\varphi_{\rm end}/2f)}\right|, \tag{1.82}$$

which is only logarithmically sensitive to $\varphi_{\rm he}$, but which can easily be large due to the condition $f \gg M_p$.

**Example III: pseudo-Goldstone dilaton**

Another case where the inflaton mass is protected by an approximate shift symmetry arises when it is a pseudo-Goldstone boson for various combinations of scaling symmetries. Because it is a scaling symmetry the same arguments that lead to trigonometric potentials for the compact phase rotations of an axionic symmetry instead in this case generically lead to exponential potentials.

In this case the form expected for the scalar potential during the inflationary regime would be

$$V = V_0 - V_1 e^{-\varphi/f} + \cdots, \tag{1.83}$$

for some scales $V_0$, $V_1$ and $f$. Our interest is in the regime $\varphi \gg f$ and in this regime $V_0$ dominates, and so is chosen as needed for inflationary cosmology, with $H_I^2 \simeq V_0/(3M_p^2)$. With this choice we have $V' \simeq (V_1/f)\, e^{-\varphi/f}$ and $V'' \simeq -(V_1/f^2)\, e^{-\varphi/f}$, leading to slow-roll parameters of the form

$$\epsilon \simeq \frac{1}{2}\left(\frac{M_p V_1}{f V_0}\right)^2 e^{-2\varphi/f} \qquad \text{and} \qquad \eta \simeq -\left(\frac{M_p^2 V_1}{f^2 V_0}\right) e^{-\varphi/f}, \tag{1.84}$$

and so $\epsilon = \frac{1}{2}(f/M_p)^2\eta^2$. Notice that these are generically small, even if $V_1 \sim V_0$, whenever $\varphi \gg f$ so there is no need to require $f$ be larger than $M_p$ to ensure a slow roll.

The number of $e$-foldings between horizon exit and $\varphi_{\rm end}$ is again given by eq. (1.75), so

$$N_e = \left(\frac{fV_0}{M_p^2 V_1}\right) \int_{\varphi_{\rm end}}^{\varphi_{\rm he}} {\rm d}\varphi \; e^{\varphi/f} = \left(\frac{f^2 V_0}{M_p^2 V_1}\right) \left[e^{\varphi_{\rm he}/f} - e^{\varphi_{\rm end}/f}\right] , \tag{1.85}$$

which can easily be large so long as $\varphi_{\rm he} \gg f$ and $\varphi_{\rm end}/f$ is order unity. It turns out that this class of models does a particularly good job of describing primordial fluctuations, and (as we shall see) the expectation that $\epsilon \sim \eta^2$ has potentially interesting observational consequences.

# 2 Cosmology: Fluctuations

This section repeats the previous discussion of ΛCDM cosmology and its peculiar initial conditions, but extends it to the properties of fluctuations about the background cosmology.

## 2.1 Structure formation in ΛCDM

Previous sections show that the universe was very homogeneous at the time of photon last scattering, since the temperature fluctuations observed in the distribution of CMB photons have an amplitude $\delta T/T \sim 10^{-5}$. On the other hand the universe around us is full of stars and galaxies and so is far from homogeneous. How did the one arise from the other?

The basic mechanism for this is based on gravitational instability: the gravitational force towards an initially over-dense region acts to attract even more material towards this region, thereby making it even more dense. This process can feed back on itself until an initially small density perturbation becomes dramatically amplified, such as into a star. This section describes the physics of this instability, in the very early universe when the density contrasts are small enough to be analyzed perturbatively in the fluctuation amplitude. The discussion follows that of ref. [9].

### 2.1.1 Nonrelativistic Density Perturbations

We start with the discussion of gravitational instability in the non-relativistic gravitating limit, both for simplicity and since this limit provides a good description of the behaviour of density fluctuations in a matter-dominated universe (which is the

one relevant for almost all of cosmology after radiation-matter decoupling occurs at $z_{\rm dec} = 1100$).

The following equations of motion describe the dynamics of a simple non-relativistic fluid with energy density, $\rho$, pressure, $p$, entropy density, $s$, and local fluid velocity $\mathbf{v}$. The equations express local conservation laws, and are

$$\begin{aligned} \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0 \qquad & \text{(energy conservation)} \\ \rho \left[ \frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla)\mathbf{v} \right] + \nabla p + \rho \nabla \phi = 0 \qquad & \text{(momentum conservation)} \\ \frac{\partial s}{\partial t} + \nabla \cdot (s \mathbf{v}) = 0 \qquad & \text{(entropy conservation)} \\ \nabla^2 \phi - 4\pi G \rho = 0 \qquad & \text{(universal gravitation)}\,, \end{aligned} \tag{2.1}$$

as well as the equation of state, $p = p(\rho, s)$. Here $\phi$ denotes the local gravitational potential.

For cosmological applications we expand about a homogeneously and radially expanding background fluid configuration. For these purposes consider a fluid background for which $\mathbf{v}_0 = H(t)\,\mathbf{r}$, where $H(t)$ is assumed a given function of $t$. In this case $\nabla \cdot \mathbf{v}_0 = 3H(t)$. This flow is motivated by the observation that it corresponds to the proper velocity if particles within the fluid were moving apart from one another according to the law $\mathbf{x}(t) = a(t)\,\mathbf{y}$, with $\mathbf{y}$ being a time-independent co-moving coordinate. In this case $\mathbf{v}_0 \equiv \mathrm{d}\mathbf{x}/\mathrm{d}t = \dot{a}\,\mathbf{y} = H(t)\,\mathbf{x}(t)$, where $H = \dot{a}/a$. In this sense $H(t)$ describes the non-relativistic analog of the Hubble parameter for the background fluid's expansion.

**Background Quantities**

We now ask what the rest of the background quantities, $\rho_0(t)$, $p_0(t)$ and $\phi_0(t)$ must satisfy in order to be consistent with this flow. The equation of energy conservation implies $\rho_0$ must satisfy

$$0 = \dot{\rho}_0 + \nabla \cdot (\rho_0 \mathbf{v}_0) = \dot{\rho}_0 + 3H\,\rho_0\,, \tag{2.2}$$

and so, given $H = \dot{a}/a$, it follows that $\rho_0 \propto a^{-3}$. That is, the non-relativistic expanding fluid necessarily requires the background density to fall with expansion as would the density in a matter-dominated universe.

Using this density in the law for universal gravitation requires the gravitational potential, $\phi_0$, take the form

$$\phi_0 = \frac{2\pi G \rho_0}{3}\,\mathbf{r}^2\,, \tag{2.3}$$

and so $\nabla\phi_0 = \frac{4}{3}\pi G\rho_0\,\mathbf{r}$. This describes the radially-directed gravitational potential which acts to decelerate the overall universal expansion.

Given this gravitational force, the momentum conservation equation, using $\dot{\mathbf{v}}_0 + (\mathbf{v}_0\cdot\nabla)\mathbf{v}_0 = [H + \dot{H}/H]\,\mathbf{v}_0$ and $\mathbf{v}_0 = H\,\mathbf{r}$, becomes

$$\left[\dot{H} + H^2 + \frac{4\pi G\rho_0}{3}\right]\mathbf{r} = 0\,. \tag{2.4}$$

This is equivalent to the Friedmann equation, as is now shown. Notice that if we take $a \propto t^\alpha$ then $H = \alpha/t$ and $\dot{H} = -\alpha/t^2 = -H^2/\alpha$. This, together with $\rho_0 \propto a^{-3} \propto t^{-3\alpha}$, is consistent with eq. (2.4) only if $\alpha = 2/3$, as expected for a matter-dominated universe. Furthermore, with this choice for $\alpha$ we also have $\dot{H} + H^2 = -\frac{1}{2}H^2$, and so eq. (2.4) is equivalent to

$$H^2 = \frac{8\pi G}{3}\,\rho_0\,, \tag{2.5}$$

which is the Friedmann equation, as claimed.

When studying perturbations we solve the entropy equation by taking $s_0 = 0$.

**Perturbations during matter domination**

To study perturbations about this background take $\mathbf{v} = \mathbf{v}_0 + \delta\mathbf{v}$, $\rho = \rho_0 + \delta\rho$, $p = p_0 + \delta p$, $s = \delta s$ and $\phi = \phi_0 + \delta\phi$, and expand the equations of motion to first order in the perturbations. Defining $D_t = \partial/\partial t + \mathbf{v}_0\cdot\nabla$, the linearized equations in this case become

$$\begin{aligned} D_t\,\delta\rho + 3H\,\delta\rho + \rho_0\nabla\cdot\delta\mathbf{v} &= 0 \\ \rho_0(D_t\,\delta\mathbf{v} + H\,\delta\mathbf{v}) + \nabla\delta p + \rho_0\nabla\delta\phi &= 0 \\ D_t\,\delta s &= 0 \\ \nabla^2\delta\phi - 4\pi G\delta\rho &= 0\,. \end{aligned} \tag{2.6}$$

To obtain this form for the momentum conservation equation requires using the equations of motion for the background quantities.

Our interest is in the evolution of $\delta\rho$, and this can be isolated by taking $D_t$ of the first of eqs. (2.6) and the divergence of the second if these equations, and using the results to eliminate $\delta\mathbf{v}$. The remaining equations involve the two basic fluid perturbations, $\delta\rho$ and $\delta s$, and imply both $D_t\,\delta s = 0$ and

$$D_t^2\left(\frac{\delta\rho}{\rho_0}\right) + 2H\,D_t\left(\frac{\delta\rho}{\rho_0}\right) - c_s^2\,\nabla^2\left(\frac{\delta\rho}{\rho_0}\right) - 4\pi G\rho_0\left(\frac{\delta\rho}{\rho_0}\right) = \frac{\xi}{\rho_0}\,\delta s\,, \tag{2.7}$$

where

$$c_s^2 := \left(\frac{\partial p}{\partial \rho}\right)_{s0} \quad \text{and} \quad \xi := \left(\frac{\partial p}{\partial s}\right)_{\rho 0} . \tag{2.8}$$

In order to analyze the solutions to this equation, it is convenient to change variables to a co-moving coordinate, $\mathbf{y}$, defined by $\mathbf{r} = a(t)\,\mathbf{y}$. In this case, for any function $f = f(\mathbf{r}, t)$ we have $(\partial f/\partial t)_{\mathbf{y}} = (\partial f/\partial t)_{\mathbf{r}} + H\mathbf{r}\cdot\nabla f = D_t\, f$, and $\nabla f = (1/a)\nabla_y f$. Fourier transforming the perturbations in co-moving coordinates, $\delta\rho/\rho_0 = \delta_k(t)\,\exp[i\mathbf{k}\cdot\mathbf{y}]$, leads to the following master equation governing density perturbations

$$\ddot{\delta}_k + 2H\,\dot{\delta}_k + \left(\frac{c_s^2\,k^2}{a^2} - 4\pi G\rho_0\right)\,\delta_k = \left(\frac{\xi}{\rho_0}\right)\,\delta s\,, \tag{2.9}$$

where the over-dot denotes $\mathrm{d}/\mathrm{d}t$.

These equations have solutions whose character depends on the relative size of $k/a$ and the Jeans wave-number,

$$k_J^2(t) = \frac{4\pi G\rho_0(t)}{c_s^2(t)} = \frac{3H^2(t)}{2\,c_s^2(t)}\,, \tag{2.10}$$

with instability occurring once $k/a \ll k_J$. Notice that so long as $c_s \sim O(1)$ the Jeans length is comparable in size to the Hubble length, $\ell_J \sim H^{-1}$. For adiabatic fluctuations ($\delta s_k = 0$) the above equation implies that the short-wavelength fluctuations ($k/a \gg k_J$) undergo damped oscillations of the form

$$\delta_k(t) \propto a^{-1/2}\,\exp\left[\pm ikc_s \int^t \frac{\mathrm{d}t'}{a(t')}\right]\,. \tag{2.11}$$

The overall prefactor of $a^{-1/2}$ shows how these oscillations are damped due to the universal expansion, or Hubble friction.

Long-wavelength adiabatic oscillations ($k/a \ll k_J$) exhibit an instability, though the background expansion dilutes the instability into a power law in $t$ rather than the exponential growth usually encountered for perturbations about a static background. This dilution occurs because the overall expansion reduces the density, and this effect fights the density increase due to gravitational collapse. The approximate solutions in this case are

$$\delta_k(t) \propto t^{2/3} \propto a(t) \qquad \text{and} \qquad \delta_k(t) \propto t^{-1} \propto a^{-3/2}(t)\,, \tag{2.12}$$

with the $\delta_k(t) \sim t^{2/3}$ solution describing the instability to gravitational collapse.

Because both the red-shifted wave-number, $k/a$, and the Jeans wave-number, $k_J$, depend on time, the overall expansion of the background can convert modes from stable

to unstable (or vice versa). Whether this conversion is towards stability or instability depends on the the time dependence of $ak_J$, which is governed by the time-dependence of the combination $aH/c_s$. If $a \propto t^\alpha$ then $aH \propto t^{\alpha-1} \propto a^{1-1/\alpha}$, and so $aH$ increases with $t$ if $\alpha > 1$ and decreases with $t$ if $\alpha < 1$. Since $\alpha = 2/3$ for the matter-dominated universe of interest here, it follows that $aH \propto t^{-1/3} \propto a^{-1/2}$, and so *decreases* with $t$. Provided that $c_s$ does not change much, this ensures that in the absence of other influences modes having fixed $k$ pass from being unstable to stable as $a$ increases due to the overall expansion.

**Perturbations during radiation and vacuum domination**

A completely relativistic treatment of density perturbations requires following fluctuations in the matter stress energy as well as in the metric itself (since these are related by Einstein's equations relating geometry and stress-energy). The details of such calculations go beyond the scope of these notes, although some of the main features are described below. But the above considerations suffice to address a result that is an important part of the structure-formation story: the stalling of Dark Matter perturbation growth during radiation- or vacuum-dominated epochs.

To contrast how fluctuations grow during radiation and matter domination it is instructive to examine the transition from radiation to matter domination. To this end we again use the above equation governing the growth of density fluctuations for non-relativistic matter,

$$\ddot{\delta}_{\mathbf{k}} + 2H\,\dot{\delta}_{\mathbf{k}} + \left(\frac{c_s^2\mathbf{k}^2}{a^2} - 4\pi G\rho_{m0}\right)\delta_{\mathbf{k}} = 0\,, \tag{2.13}$$

where $H^2 = 8\pi G\rho_0/3$ and $\rho_0 = \rho_{m0} + \rho_{r0}$ is no longer the same as $\rho_{m0}$. During the transition between radiation and matter domination, we use

$$H^2(a) = \frac{8\pi G\rho_0}{3} = \frac{H_{\rm eq}^2}{2}\left[\left(\frac{a_{\rm eq}}{a}\right)^3 + \left(\frac{a_{\rm eq}}{a}\right)^4\right], \tag{2.14}$$

where radiation-matter equality occurs when $a = a_{\rm eq}$, at which point $H(a = a_{\rm eq}) = H_{\rm eq}$. As described around eq. (2.4), any departure from the choice $a(t) \propto t^{2/3}$ — such as occurs when radiation dominates in $\rho(a)$ — precludes solving background momentum-conservation equation, but this does not present a problem because (2.4) can instead be replaced by the full radiation-dominated Friedmann equation without changing the description of the response of the nonrelativistic fluctuations.

For all modes for which the pressure term, $c_s^2\,\mathbf{k}^2/a^2$, is negligible, $\delta(x)$ satisfies

$$2x(1+x)\,\delta'' + (3x+2)\,\delta' - 3\,\delta = 0\,, \tag{2.15}$$

where the scale factor, $x = a/a_{\rm eq}$, is used as a proxy for time and primes denote differentiation with respect to $x$. As is easily checked, this is solved by $\delta^{(1)} \propto \left(x + \frac{2}{3}\right)$, and so the growing mode during matter domination does not also grow during radiation domination. Use of the Frobenius method shows that a linearly independent solution behaves for $x \ll 1$ (*i.e.* deep in the radiation-dominated regime) as $\delta^{(2)} \propto \delta^{(1)} \ln x +$ (analytic) where 'analytic' denotes a simple power series proportional to $1 + c_1 x + \cdots$. These solutions show how density perturbations for non-relativistic matter grow at most logarithmically during the radiation-dominated epoch.

A similar analysis covers the case where Dark Energy (modelled as a cosmological constant) dominates in an $\Omega = 1$ universe. In this case $4\pi G\rho_{m0} \sim \Omega_m H^2 \ll H^2$ and so the instability term becomes negligible relative to the first two terms of (2.13). This leads to

$$\ddot{\delta} + 2H\,\dot{\delta} \simeq 0\,, \tag{2.16}$$

which has as solution $\dot{\delta} \propto a^{-2}$. Integrating again gives a frozen mode, $\delta \propto a^0$, and a damped mode that falls as $\delta \propto a^{-2}$ when $H$ is constant (as it is when Dark Energy dominates and $a \propto e^{Ht}$). This shows that non-relativistic density perturbations stop growing again once matter domination ends.

We are now in a position to summarize how inhomogeneities grow in the late universe, assuming the presence of an initial spectrum of very small primordial density fluctuations. The key observation is that several conditions all have to hold in order for there to be appreciable growth of density inhomogeneities. These conditions are:

1. No fluctuations grow appreciably at all unless the Universe is matter dominated.

2. Fluctuations of any type do not grow for super-Hubble modes, for which $k/a \ll H$, regardless of what type of matter dominates the background evolution.

3. Nonrelativistic matter in a matter-dominated universe are unstable, but only for those modes in the momentum range $H \ll (k/a) \ll H/c_s$, and these grow proportional to the scale factor: $\delta_k \propto a$.

Before pursuing the implications of these conditions for instability, we pause to describe what properties of fluctuations are actually measured.

### 2.1.2 The Power Spectrum

The presence of unstable density fluctuations implies the universe does not remain precisely homogeneous and isotropic once matter domination begins, and so the view

seen by observers like us depends on their locations in the universe relative to the fluctuations. For this reason, when comparing with observations it is less useful to try to track the detailed form of a specific fluctuation and instead better to characterize fluctuations by their statistical properties, since these can be more directly applied to observers without knowing their specific place in the universe. In particular we imagine there being an ensemble of density fluctuations, whose phases we assume to be uncorrelated and whose amplitudes are taken to be random variables.

On the observation side statistical inferences can be made about the probability distribution governing the distribution of fluctuation amplitudes by measuring statistical properties of the matter distribution observed around us. For instance, a useful statistic measures the mass-mass auto-correlation function

$$\xi(\mathbf{r}-\mathbf{r}') \equiv \frac{\langle \delta\rho(\mathbf{r})\, \delta\rho(\mathbf{r}')\rangle}{\langle\rho\rangle^2}\,, \tag{2.17}$$

which might be measured by performing surveys of the positions of large samples of galaxies.[10] When using (2.17) with observations the average $\langle\cdots\rangle$ is interpreted as integration of one of the positions (say, $\mathbf{r}'$) over all directions in the sky.[11]

When making predictions $\langle\cdots\rangle$ instead is regarded as an average over whatever ensemble is thought to govern the statistics of the fluctuations $\delta_k$. Fourier transforming $\delta\rho(\mathbf{r})/\langle\rho\rangle = \int \mathrm{d}^3k\, \delta_k\, \exp[i\mathbf{k}\cdot\mathbf{r}]$ in comoving coordinates, as before, allows $\xi(\mathbf{r})$ to be related to the following ensemble average over the Fourier mode amplitudes, $\delta_k$.

$$\xi(r) = \int \frac{\mathrm{d}^3k}{(2\pi)^3}\, \langle|\delta_k|^2\rangle\, \exp[i\mathbf{k}\cdot\mathbf{r}] = \frac{1}{2\pi^2}\int_0^\infty \frac{\mathrm{d}k}{k}\, k^3\, P_\rho(k) \left(\frac{\sin kr}{kr}\right)\,, \tag{2.18}$$

which defines the density *power spectrum*: $P_\rho(k) := \langle|\delta_k|^2\rangle$.

For homogeneous and isotropic backgrounds $P_\rho(k)$ depends only on the magnitude $k = |\mathbf{k}|$ and not on direction, and this is used above to perform the angular integrations. The average in these expressions is over the ensemble, and it is this average which collapses the right-hand side down to a single Fourier integral. The last equality motivates the definition

$$\Delta^2_\rho(k) := \frac{k^3}{2\pi^2}\, P_\rho(k)\,, \tag{2.19}$$

[10] A practical complication arises because although galaxies are relatively easy to count, most of the mass density is actually Dark Matter. Consequently assumptions are required to relate these to one another; the usual choice being that the galaxy and mass density functions are related to one another through a phenomenologically defined 'bias' factor.

[11] The density correlation function can also be measured using the temperature fluctuations of the CMB, because these fluctuations can be interpreted as redshifts acquired by CMB photons as they climb out of the gravitational potential wells formed by density fluctuations in nonrelativistic matter.

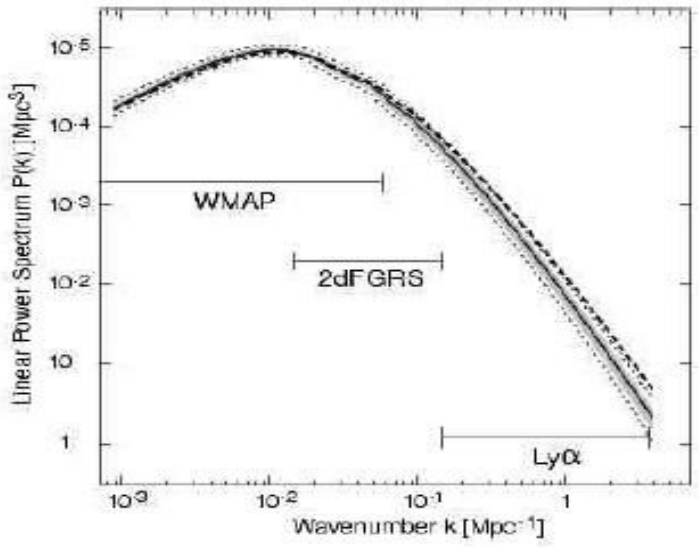


**Figure 5**. The power spectrum as obtained from WMAP measurements of the CMB spectrum, together with the 2dF Galaxy Redshift Survey and Lyman $\alpha$ measurements.

since this quantity would be expected to be independent of $k$ if the distribution of described by $P_\rho(k)$ were scale invariant.

A variety of observations over the years give the form of $P_\rho(k)$ as inferred from the distribution of structure around us, with results summarized in Figure 5. As illustrated in Figure 6 the scale $k$ appearing in $P_\rho(k)$ is correlated with how far back one looks into the universe, with measurements of distant objects in the remote past determining the shape of $P_\rho(k)$ for small $k$, and measurements of more nearby objects in the more recent past constraining $P_\rho(k)$ for larger $k$. As indicated in Figure 5 inferences about the shape of $P_\rho(k)$ for small $k$ come from measurements of the temperature fluctuations in the CMB; those at intermediate $k$ come from galaxy distributions as obtained through galaxy surveys and those at the largest $k$ come from measurements of the how quasar light is absorbed by intervening Hydrogen gas clouds, the so-called Lyman-$\alpha$ 'forest'.

These observations are well approximated by the phenomenological formula,

$$P(k) = \frac{A\,k^{n_s}}{(1+\alpha\,k+\beta\,k^2)^2}\,, \tag{2.20}$$

where

$$\alpha = 16\left(\frac{0.5}{\Omega\,h^2}\right)\ \text{Mpc} \quad \text{and} \quad \beta = 19\left(\frac{0.5}{\Omega\,h^2}\right)^2\ \text{Mpc}^2 \quad \text{and} \quad n_s = 0.96\,. \tag{2.21}$$

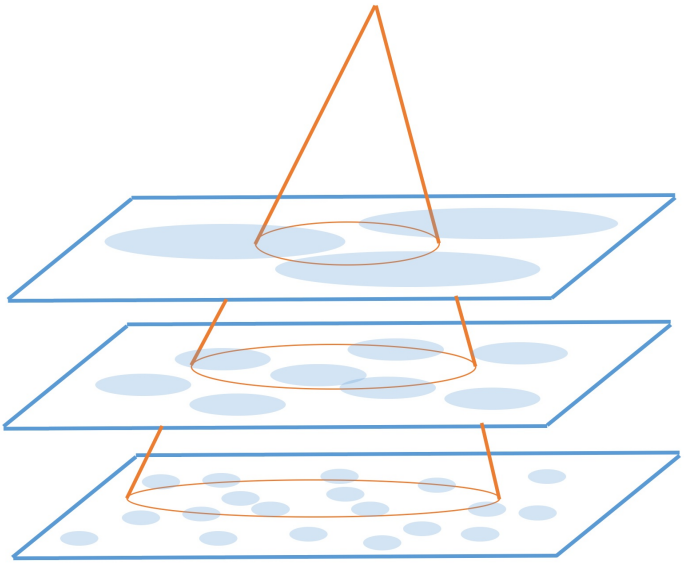

**Figure 6**. A sketch of several spatial slices intersecting the past light cone of an astronomer on Earth. The orange ovals indicate how the light cone has larger intersections with the spatial slices the further back one looks. The pale blue ovals indicate regions the size of the Hubble distance on each spatial slice. Correlations outside of these ovals (such as the uniformity of the CMB temperature) represent a puzzle for ΛCDM cosmology. The figure shows how later times (higher slices) have larger Hubble distances, as well as how observations only sample the largest distance scales on the most remote spatial slices. This illustrates why CMB measurements tend to constrain the power spectrum for small $k$ while observations of more nearby objects (like galaxy distributions or the distribution of foreground Lyman-$\alpha$ Hydrogen gas clouds) constrain larger $k$.

Here $h = H_0/(100 \text{ km/sec/Mpc}) \approx 0.7$, and $\Omega \approx 1$ denotes the present value of $\rho/\rho_c$. Given that $n_s \approx 1$ the observations suggest the power spectrum is close to linear, $P(k) \propto k$ for $k \ll k_\star \sim 0.07$ Mpc$^{-1}$, and $P(k) \propto k^{-3}$ for $k \gg k_\star$. The value $k_\star$ here is simply defined to be the place where $P_\rho(k)$ turns over and makes the transition from $P_\rho \propto k$ to $P_\rho \propto k^{-3}$.

As described below, there are good reasons to believe that the shape of $P_\rho(k)$ for $k \ll k_\star$ represents the pattern of primordial fluctuations inherited from the very early universe, while the shape for $k > k_\star$ reflects how fluctuations evolve in the later universe. Consequently observations are consistent with primordial fluctuations being

close to[12] a scale-invariant *Zel'dovich* spectrum, $P_\rho(k) = A\,k$, corresponding to $n_s = 1$. As we also see below, the result $n_s$ is predicted to be close to, but not equal to, unity by inflationary models.

For later purposes it proves more convenient to work with the power spectrum for the Newtonian gravitational potential, $\delta\phi$, that is related to $\delta\rho$ by Poisson's equation — the last of eqs. (2.6) — and so $\delta\phi_k \propto \delta_k/k^2$. Because of this relation their power spectra are related by $P_\phi(k) = P_\rho(k)/k^4$ as well as

$$\Delta^2_\phi(k) := \frac{k^3}{2\pi^2}\,P_\phi(k) = \frac{P_\rho(k)}{2\pi^2 k} \propto \begin{cases} k^{n_s-1} \text{ if } & k \ll k_\star \\ k^{n_s-5} \text{ if } & k \gg k_\star \end{cases} . \tag{2.22}$$

This last expression also clarifies why the choice $n_s = 1$ is called scale invariant. When $n_s = 1$ the primordial ($k \ll k_\star$) spectrum for $\Delta^2_\phi(k)$ becomes $k$-independent, as would be expected for a scale-invariant process.

### 2.1.3 Late-time structure growth

Before trying to explain the properties of the primordial part of the power spectrum — $\Delta^2_\phi(k) \propto Ak^{n_s-1}$ — we first digress to explain the explanation for why the measured distribution has the peculiar hump-shaped form, bending at $k \simeq k_\star$. This shape arises due to the processing of density fluctuations by the evolution of Dark Matter in the subsequent universe, as we now describe.

The key observations go back to the three criteria, given at the end of §2.1.1, for when fluctuating modes can grow. These state that the fluctuations that are most important are those involving nonrelativistic matter, although these remain frozen unless the universe is matter dominated and the mode number lies within the interval $H \ll k/a \ll H/c_s$. These conditions for growth superimpose a $k$-dependence on $P_\rho(k)$, for the following reasons.

The important wave-number $k_\star$ corresponds to the wave-number, $k_{\rm eq}$, for which modes satisfy $k/a \sim H$ at the epoch of radiation-matter equality (which occurs at $z_{\rm eq} = 3600$). Numerically, $k_{\rm eq}$ corresponds to a co-moving wave-number of order $k_{\rm eq} \sim 0.07\ {\rm Mpc}^{-1}$. What is important about this scale is that it divides modes (with $k > k_{\rm eq}$) that re-enter the Hubble scale during radiation domination and those (with $k < k_{\rm eq}$) that re-enter during matter domination.

Because they re-enter during matter domination, all Dark Matter fluctuation modes with $k < k_{\rm eq}$ are free to begin growing immediately on re-entry and have done so ever

[12] Close to but not equal to. Fits to ΛCDM cosmology establish $n_s$ is significantly different from 1.

since, at least until the very recent advent of Dark Energy domination.[13] So the present-day power spectrum for these modes reflects the primordial one which was frozen into these modes long ago when they left the Hubble scale in the pre-ΛCDM era. It is these modes that reveal the primordial distribution

$$P(k) \propto k^{n_s} \qquad (\text{for } k \ll k_{\rm eq})\,. \tag{2.23}$$

By contrast, those modes with $k \gg k_{\rm eq}$ re-enter the Hubble scale during the radiation-dominated epoch that precedes matter-radiation equality. The amplitude of these modes therefore remain frozen at their values at the time of re-entry, because they are unable to grow while the Universe is radiation dominated. Consequently they remain stunted in amplitude relative to their longer-wavelength counterparts while waiting for the matter to become matter-dominated, leading to a suppression of $P_\rho(k)$ for $k \gg k_{\rm eq}$.

The *relative* stunting of large-$k$ modes relative to small-$k$ modes can be computed from the information that the unstable modes grow with amplitude $\delta_k(a) \propto a$ during matter-domination. For $k < k_{\rm eq}$ this growth applies as soon as they cross the Hubble scale, while for $k > k_{\rm eq}$ the modes cannot grow in this way until the transition from radiation to matter domination. As a result the relative size of two modes, one with $k_0 \ll k_{\rm eq}$ and one with $k \gg k_{\rm eq}$, is

$$\frac{\delta_k(a)}{\delta_{k_0}(a)} \propto \frac{\delta_k(a_k)(a/a_{\rm eq})}{\delta_{k_0}(a_{k_0})(a/a_{k_0})} \propto \frac{\delta_k(a_k)(a/a_k)}{\delta_{k_0}(a_{k_0})(a/a_{k_0})} \left(\frac{k_{\rm eq}}{k}\right)^2\,, \tag{2.24}$$

where $a_k$ denotes the scale factor at the ($k$-dependent) epoch of re-entry, defined by $k = a_k H_k$. The first relation in (2.24) uses that modes in the numerator all start growing at the same time (radiation-matter equality), while those in the denominator grow for a $k_0$-dependent amount $a/a_{k_0}$. The second relation then makes the $k$-dependence of the suppression $a_k/a_{\rm eq}$ in the numerator explicit, using the matter-domination evolution $aH \propto a^{-1/2}$ in the re-entry condition to conclude $k = a_k H_k \propto a_k^{-1/2}$ and so $a_k \propto k^{-2}$.

This leads to the expectation that the power spectrum has the form $P(k) = P_{\rm prim}(k)\,\mathcal{T}(k)$, where $P_{\rm prim}(k) = \langle|\delta_k(a)|^2\rangle = \langle|\delta_k(a_k)|^2\rangle(a/a_k)^2$ is the primordial power spectrum and $\mathcal{T}(k)$ is the transfer function that expresses the relative stunting of modes

[13] For most modes $\delta_k \simeq \mathcal{O}(1)$ occurs before Dark Energy domination, at which point nonlinear gravitational physics is expected to produce the large-scale structure actually seen in galaxy surveys. It is noteworthy that there would not have been sufficient time for modes small enough to describe the CMB to become nonlinear if baryons were the only non-relativistic matter present, and this is part of the evidence for Dark Matter's existence.

for $k \gg k_{\rm eq}$. Keeping in mind that $P(k) \propto |\delta_k|^2$ the above discussion shows we expect $\mathcal{T}(k) \simeq 1$ for $k \ll k_{\rm eq}$ and $\mathcal{T}(k) \simeq (k_{\rm eq}/k)^4$ for $k \gg k_{\rm eq}$. Given a primordial distribution $P_{\rm prim}(k) \simeq Ak^{n_s}$ this leads to

$$P_\rho(k) \propto \begin{cases} k^{n_s} & \text{if} \quad k \ll k_\star \\ k^{n_s-4} & \text{if} \quad k \gg k_\star \end{cases}, \tag{2.25}$$

much as is observed.

### 2.2 Primordial fluctuations from inflation

The previous discussion shows that fluctuations in the ΛCDM model also provide a successful description of structure in the universe, but only given the initial condition of a primordial spectrum of fluctuations having a specific power-law form: $P_\rho(k) \simeq A_s k^{n_s}$ (or $\Delta^2_\phi(k) \simeq A_s k^{n_s-1}$). It again falls to the earlier universe to explain why primordial fluctuations should have this specific form, and why it should be robust against the many poorly understood details governing the physics of this earlier epoch.

It is remarkable that there is evidence that an earlier period of inflationary expansion can also explain this initial distribution of fluctuations. This section provides a sketch of this evidence. Since the modes of interest start off during ΛCDM outside the Hubble length, $k \ll aH$, and are known to be small, their evolution can be tracked into earlier epochs using linear perturbation theory. Because the modes are super-Hubble in size the treatment must be relativistic, and so involves linearizing the coupled Einstein-matter field equations. The first part of this section sketches how this super-Hubble evolution works, and shows how to relate the primordial fluctuations that re-enter the Hubble scale to those that exit the Hubble scale during the inflationary epoch (see Figure 4).

At first sight this just pushes to problem back to an earlier time, requiring an explanation why a particular pattern of fluctuations should exist during inflation. Even worse, within the classical approximation there is good reason to believe there should be no fluctuations at all leaving at horizon exit during inflation. This is because the exponential growth of the scale factor, $a \propto e^{Ht}$, during inflation is absolutely ruthless in ironing out any spacetime wrinkles since momentum-dependent terms like $(k/a)^2$ in the field equations go to zero so quickly.

But the key words in the above are "within the classical approximation". Quantum fluctuations are *not* ironed away during inflation, and persist at a level proportional to the Hubble scale. Because this Hubble scale is approximately constant the resulting fluctuations are largely scale-independent, providing a natural explanation for why

primordial fluctuations seem to be close to the Zel'dovich spectrum. But $H$ during inflation also cannot be exactly constant since inflation must end eventually. In the explicit models examined earlier the time-dependence of $H$ arises at a level suppressed by the slow-roll parameters $\epsilon$ and $\eta$ and so deviations from scale invariance should arise at the few percent level. Because of this we shall find below that the prediction for $n_s$ in inflationary models is a bit smaller than unity, naturally agreeing with the observed value $n_s \simeq 0.96$.

### 2.2.1 Linear evolution of metric-inflaton fluctuations

The first task is to evolve fluctuations forward from the epoch of inflationary horizon exit until they re-enter during the later Hot Big Bang era. In particular our focus is on the perturbations of the metric, $\delta g_{\mu\nu}$, since these include perturbations of the Newtonian potential and so also the density fluctuations whose power spectrum is ultimately measured. The discussion here follows that of [9].

The symmetry of the FRW background allows the fluctuations of the metric to be classified by their rotational properties, with fluctuations of different spin not mixing at linear order in the field equations. Fluctuations of the metric come in three such kinds: *scalar*, *vector* and *tensor* fluctuations. Specializing to a spatially flat FRW background and transforming to conformal time, $\tau = \int \mathrm{d}t/a$, the scalar perturbations may be written

$$\delta_S g_{\mu\nu} = a^2 \begin{pmatrix} 2\phi & \partial_j \mathcal{B} \\ \partial_i \mathcal{B} & 2\psi\, \delta_{ij} + \partial_i \partial_j \mathcal{E} \end{pmatrix} , \tag{2.26}$$

while the vector and tensor ones are

$$\delta_V g_{\mu\nu} = a^2 \begin{pmatrix} 0 & \mathcal{V}_j \\ \mathcal{V}_i & \partial_i \mathcal{W}_j + \partial_j \mathcal{W}_i \end{pmatrix} \quad \text{and} \quad \delta_T g_{\mu\nu} = a^2 \begin{pmatrix} 0 & 0 \\ 0 & h_{ij} \end{pmatrix} . \tag{2.27}$$

Here all vectors and tensors are divergence-free, as is the tensor (which is also traceless). To this is to be added the fluctuations in the inflaton field, $\varphi(t) + \delta\varphi$.

There is great freedom to modify these functions by performing infinitesimal coordinate transformations, so it is useful to define the following combinations that are invariant at linearized order:

$$\begin{aligned} \Phi = \phi - \frac{1}{a}\Big[a(\mathcal{B} - \mathcal{E}')\Big]' , \qquad \Psi = \psi + \frac{a'}{a}(\mathcal{B} - \mathcal{E}') \\ \delta\chi = \delta\varphi - \varphi'(\mathcal{B} - \mathcal{E}') , \quad V_i = \mathcal{V}_i - \mathcal{W}_i \quad \text{and} \quad h_{ij} , \end{aligned} \tag{2.28}$$

in terms of which all physical inferences can be drawn. Here primes denote differentiation with respect to conformal time, $\tau$. Notice that $\Phi$, $\Psi$ and $V_i$ reduce to $\phi$, $\psi$ and $\mathcal{V}_i$

in the gauge choice where $\mathcal{B} = \mathcal{E} = \mathcal{W}_i = 0$, and so $\Phi$ is the relativistic generalization of the Newtonian potential.

These functions are evolved forward in time by linearizing the relevant field equations:

$$\Box\varphi - V'(\varphi) = 0 \quad \text{and} \quad R_{\mu\nu} - \frac{1}{2}\, R g_{\mu\nu} = \frac{T_{\mu\nu}}{M_p^2}\,, \tag{2.29}$$

and provided we use the invariant stress-energy perturbations,

$$\begin{aligned}
\delta\mathcal{T}^0{}_0 &= \delta T^0{}_0 - \left[t^0{}_0\right]'(\mathcal{B} - \mathcal{E}')\,,\\
\delta\mathcal{T}^0{}_i &= \delta T^0{}_i - \left[t^0{}_0 - \frac{1}{3}\, t^k{}_k\right]\partial_i(\mathcal{B} - \mathcal{E}')\,,\\
\delta\mathcal{T}^i{}_j &= \delta T^i{}_j - \left[t^i{}_j\right]'(\mathcal{B} - \mathcal{E}')\,,
\end{aligned} \tag{2.30}$$

(where $t^\mu{}_\nu$ denotes the background stress-energy), the results can be expressed purely in terms of the gauge-invariant quantities, eqs. (2.28).

The equations which result show that in the absence of vector stress-energy perturbations (*i.e.* if $\delta\mathcal{T}^0{}_i$ is a pure gradient - as would be the case for perturbed inflaton), then vector perturbations, $V_i$, are not sourced, and decay very rapidly in an expanding universe, allowing them to be henceforth ignored. Similarly, in the absence of off-diagonal stress-energy perturbations (*i.e.* if $\delta\mathcal{T}^i{}_j = \delta p\, \delta^i_j$) it is also generic that $\Psi = \Phi$.

Switching back to FRW time, the equations which govern the evolution of tensor modes then become (after Fourier transforming)

$$\ddot{h}_{ij} + 3H\,\dot{h}_{ij} + \frac{k^2}{a^2}\, h_{ij} = 0\,, \tag{2.31}$$

showing that these evolve independent of all other fluctuations. Such primordial tensor fluctuations can be observable if they survive into the later universe, since the differential stretching of spacetime that they predict can contribute observably to the polarization of the CMB photons. The search for evidence for this type of primordial tensor fluctuations is active and ongoing, and we shall see is expected in inflation to be characterized by a near scale-invariant tensor power spectrum,

$$P_h(k) \propto A_T\, k^{n_T}\,. \tag{2.32}$$

The equations evolving the scalar fluctuations are more complicated and similarly

reduce to

$$\delta\ddot{\chi} + 3H\delta\dot{\chi} + \frac{k^2}{a^2}\delta\chi + V''(\varphi)\delta\chi - 4\dot{\varphi}\,\dot{\Phi} + 2V'(\varphi)\,\Phi = 0$$
$$\text{and} \qquad \dot{\Phi} + H\,\Phi = \frac{\dot{\varphi}}{2M_p^2}\,\delta\chi\,. \tag{2.33}$$

The homogeneous background fields themselves satisfy the equations

$$\ddot{\varphi} + 3H\dot{\varphi} + V'(\varphi) = 0 \quad \text{and} \quad 3M_p^2H^2 = \frac{1}{2}\dot{\varphi}^2 + V(\varphi)\,. \tag{2.34}$$

These expressions show that although $\Phi$ and $\delta\chi$ would decouple from one another if expanded about a static background (for which $\dot{\varphi} = V' = 0$), they do not when the background is time-dependent.

### 2.2.2 Slow-roll evolution of scalar perturbations

The character of the solutions of these equations depends strongly on the size of $k/a$ relative to $H$, since this dictates the extent to which the frictional terms can compete with the spatial derivatives. As usual the two independent solutions for $\delta\chi$ that apply when $k/a \gg H$ describe damped oscillations

$$\delta\chi_k \propto \frac{1}{a\sqrt{k}} \exp\left[\pm ik \int^t \frac{\mathrm{d}t'}{a(t')}\right]. \tag{2.35}$$

Our interest during inflation is in the limit $k/a \ll H$ in a slow-roll regime for which $\delta\ddot{\chi}$, $\ddot{\varphi}$ and $\dot{\Phi}$ can be neglected. In this limit the scalar evolution equations simplify to

$$3H\delta\dot{\chi} + V''(\varphi)\delta\chi + 2V'(\varphi)\Phi \simeq 0 \quad \text{and} \quad 2M_p^2H\,\Phi \simeq \dot{\varphi}\,\delta\chi\,, \tag{2.36}$$

and have approximate solutions (after Fourier transformation) of the form

$$\delta\chi_k \simeq C_k\,\frac{V'(\varphi)}{V(\varphi)} \quad \text{and} \quad \Phi_k \simeq -\frac{C_k}{2}\left(\frac{V'(\varphi)}{V(\varphi)}\right)^2. \tag{2.37}$$

where $C_k$ is a (potentially $k$-dependent) constant of integration. Since the background fields satisfy $M_pV'/V = \sqrt{2\epsilon}$ these equations show how the amplitude of $\delta\chi_k$ and $\Phi_k$ during inflation track the evolution of the slow-roll parameter, $\epsilon$, for super-Hubble modes, and therefore tend to grow in amplitude as inflation eventually draws to a close.

We have two remaining problems: (*i*) What is the origin of the initial fluctuations at horizon exit? (*ii*) How do we evolve fluctuations from the end of inflation through to the later epoch of horizon re-entry? The latter of these seems particularly vicious since it *a priori* might be expected to depend on the many details involved in getting the Universe from its inflationary epoch to the later Hot Big Bang.

### 2.2.3 Post-Inflationary evolution

For the case of single-field inflation discussed here, post-inflationary evolution of the fluctuation $\Phi$ actually turns out to be quite simple. This is because it can be shown that when $k \ll aH$ the quantity

$$\zeta = \Phi + \frac{2}{3}\left(\frac{\Phi + \dot{\Phi}/H}{1+w}\right) = \frac{1}{3(1+w)}\left[(5+3w)\,\Phi + \frac{2\dot{\Phi}}{H}\right], \tag{2.38}$$

is *conserved*, $\dot{\zeta} \simeq 0$. This result has been proven under a wide variety of assumptions [10], but the form we use here assumes that the background cosmology satisfies an equation of state $p = w\rho$, but $w$ is *not* assumed to be constant. The same result is known not to be true if there were more than a single scalar field evolving.

Conservation of $\zeta$ is a very powerful result because it can be used to evolve fluctuations using $\zeta(t_i) = \zeta(t_f)$, assuming only that they involve a single scalar field, and that the modes in question are well outside the horizon: $k/a \ll H$. Furthermore, although $\dot{\Phi}$ in general becomes nonzero at places where $w$ varies strongly with time, this time dependence quickly damps due to Hubble friction for modes outside the Hubble scale.

We may therefore for most of the universe's history also neglect the dependence of $\zeta$ on $\dot{\Phi}$ provided we restrict $t_i$ and $t_f$ to epochs during which $w$ is roughly constant. This allows the expression $\zeta(t_i) = \zeta(t_f)$ to be simplified to

$$\Phi_f = \frac{1+w_f}{1+w_i}\left(\frac{5+3w_i}{5+3w_f}\right)\Phi_i\,, \tag{2.39}$$

where $w_i = w(t_i)$ and $w_f = w(t_f)$, implying in particular $\Phi_f = \Phi_i$ whenever $w_i = w_f$. Similarly, the values of $\Phi$ deep within radiation and matter dominated phases are related by $\Phi_{\rm mat} \simeq \frac{9}{10}\,\Phi_{\rm rad}$.

To infer the value of $\Phi$ in the later Hot Big Bang era we choose $t_i$ just after horizon exit (where a simple calculation shows $w_i \simeq -1 + \frac{2}{3}\,\epsilon_{\rm he}$, with $\epsilon_{\rm he}$ the slow-roll parameter at horizon exit). $t_f$ is then chosen in the radiation dominated universe (where $w_f = \frac{1}{3}$), either just before horizon re-entry for the mode of interest, or just before the transition to matter domination, whichever comes first. Eqs. (2.37) and (2.39) then imply

$$\Phi_f \simeq \left(\frac{6\,\Phi}{\epsilon}\right)_{\rm he}. \tag{2.40}$$

It remains to grapple with what should be expected for the initial condition for $\Phi$ at horizon exit.

### 2.2.4 Quantum origin of fluctuations

The primordial fluctuation amplitude derived in this way depends on the integration constants $C_k$, which are themselves set by the initial conditions for the fluctuation at horizon exit, during inflation. But why should this amplitude be nonzero given that all previous evolution is strongly damped, as in eq. (2.35)? The result remains nonzero (and largely independent of the details of earlier evolution) because quantum fluctuations in $\delta\chi$ continually replenish the perturbations long after any initial classical configurations have damped away.

The starting point for the calculation of the amplitude of scalar perturbations is the observation that the inflaton and metric fields whose dynamics we are following are quantum fields, not classical ones. For instance, for spatially-flat spacetimes the linearized inflaton field, $\delta\chi$, is described by the operator

$$\delta\chi(x) = \int \frac{\mathrm{d}^3k}{(2\pi)^3}\left[c_k\, u_k(t)\, e^{i\mathbf{k}\cdot\mathbf{r}/a} + c_k^*\, u_k^*(t)\, e^{-i\mathbf{k}\cdot\mathbf{r}/a}\right], \tag{2.41}$$

where we expand in a basis of eigenmodes of the scalar field equation in the background metric, $u_k(t)\, e^{i\mathbf{k}\cdot\mathbf{x}}$, labelled by the co-moving momentum $\mathbf{k}$. For constant $H$ the time-dependent mode functions are

$$u_k(t) \propto \frac{H}{k^{3/2}}\left(i + \frac{k}{aH}\right)\exp\left(\frac{ik}{aH}\right), \tag{2.42}$$

which reduces to the standard flat-space form (up to a slowly-varying phase), $u_k(t) \propto a^{-1}k^{-1/2}\, e^{-ikt/a}$, when $k/a \gg H$. (This is perhaps easiest to see using conformal time, for which $\exp(ik/aH) = \exp(-ik\tau)$.) The quantities $c_k$ and their adjoints $c_k^*$ are *annihilation* and *creation operators*, which define the adiabatic vacuum state, $|\Omega\rangle$, through the condition $c_k|\Omega\rangle = 0$ (for all $\mathbf{k}$).

The $\delta\chi$ auto-correlation function in this vacuum, $\langle\delta\chi(x)\delta\chi(x')\rangle$, describes the quantum fluctuations of the field amplitude in the quantum ground state, and the key assumption is that the quantum statistics of the mode leaving the horizon during inflation agrees with the classical fluctuations of the field $\delta\chi$ after evolving outside of the Hubble scale. This assumes the quantum fluctuations to be decohered (for preliminary discussions see ref. [11, 12]) into classical distribution for $\delta\chi$ sometime between horizon exit and horizon re-entry.

It turns out that during inflation interactions with the bath of short-wavelength, sub-Hubble modes is extremely efficient at decohering the quantum fluctuations of long-wavelength, super-Hubble modes [13]. As is usual when a system is decohered through

interactions with an environment, the resulting classical distribution is normally defined for the 'pointer basis', that diagonalizes the interactions with the environment. It turns out that the freezing of super-Hubble modes has the effect of making them very classical (WKB-like), and so ensure the fields canonical momenta become functions of the fields themselves. This ensures that it is always the field basis that diagonalizes any local interactions, and so guarantees that quantum fluctuations become classical fluctuations for the fields (like $\delta\chi$) rather than (say) their canonical momenta.

The upshot is that after several $e$-foldings even very weak interactions (like gravitational strength ones) eventually convert quantum fluctuations into classical statistical fluctuations for the classical field, $\varphi$, about its spatial mean. For practical purposes, this means in the above calculations we can simply use the initial condition $|\delta\chi_k| \sim [\langle \delta\chi_k \delta\chi_{-k}\rangle]^{1/2} \propto |u_k(t)|$. For observational purposes what matters is that the classical variance of these statistical fluctuations is well-described by the corresponding quantum auto-correlations – a property that relies on the kinds of 'squeezed' quantum states that arise during inflation [9, 14].

Evaluating $\delta\chi_k \sim u_k$ at $t_{\rm he}$ (where $k = aH$) and equating the result to the fluctuation of eq. (2.37) allows the integration constant in this equation to be determined to be

$$C_k = u_k(t_{\rm he}) \left(\frac{V}{V'}\right)_{\varphi_{\rm he}} , \tag{2.43}$$

where both $t_{\rm he}$ and $\varphi_{\rm he} = \varphi(t_{\rm he})$ implicitly depend on $k$. Using this to compute $\Phi_k$ in eq. (2.37) then gives

$$\Phi_k(t) = -\frac{1}{2} u_k(t_{\rm he}) \left(\frac{V}{V'}\right)_{\varphi_{\rm he}} \left(\frac{V'}{V}\right)^2_{\varphi(t)} = -\epsilon(t) \left(\frac{u_k}{\sqrt{2\epsilon}\, M_p}\right)_{t_{\rm he}} . \tag{2.44}$$

In particular, evaluating at $t = t_{\rm he}$ then gives

$$\Phi_k(t_{\rm he}) = -\left(\frac{u_k}{M_p}\sqrt{\frac{\epsilon}{2}}\right)_{t_{\rm he}} . \tag{2.45}$$

### 2.2.5 Predictions for the scalar power spectrum

We are now in a situation to pull everything together and compute in more detail the inflationary prediction for the properties of the primordial fluctuation spectrum. Using (2.45) in (2.40) gives

$$\Phi_k(t_f) \simeq \left(\frac{6\,\Phi}{\epsilon}\right)_{\rm he} = -\left(\frac{6u_k}{\sqrt{2\epsilon}\, M_p}\right)_{t_{\rm he}} . \tag{2.46}$$

Using this in the definition of the dimensionless power spectrum for $\Phi$, $\Delta^2_\Phi = k^3 P_\Phi/(2\pi^2)$, then leads to

$$\Delta^2_\Phi(k) \sim k^3 |\Phi_k(t_f)|^2 \sim \frac{|k^{3/2} u_k(t_{\rm he})|^2}{\epsilon(\varphi_{\rm he}) M_p^2} \sim \left(\frac{H^2}{\epsilon M_p^2}\right)_{\varphi_{\rm he}} \sim \left(\frac{V}{\epsilon M_p^4}\right)_{\varphi_{\rm he}} . \tag{2.47}$$

Once the order-unity factors are included one finds

$$\Delta^2_\Phi(k) = \frac{k^3 P_\Phi(k)}{2\pi^2} = \left(\frac{H^2}{8\pi^2 M_p^2\, \epsilon}\right)_{\rm he} = \left(\frac{V}{24\pi^2 M_p^4\, \epsilon}\right)_{\rm he} , \tag{2.48}$$

It is the quantity $V/\epsilon$ that controls the amplitude of density fluctuations, and so is to be compared with the observed power spectrum of scalar density fluctuations,

$$\Delta^2_\Phi(\hat{k}) = 2.28 \times 10^{-9} , \tag{2.49}$$

when evaluated at the reference 'pivot' point $k = \hat{k} \sim 7.5 a_0 H_0$. In terms of $V$ this implies

$$\left(\frac{V}{\epsilon}\right)^{1/4} = 6.6 \times 10^{16}\ {\rm GeV} , \tag{2.50}$$

and the smaller $\epsilon$ becomes, the smaller a potential energy during inflation is required. For $\epsilon \sim 0.01$ we have $V \sim 2 \times 10^{15}$ GeV. This is titillatingly close to the scale where the couplings of the three known interactions would unify in Grand Unified models, which may indicate a connection between the physics of Grand Unification and inflation.[14]

Notice also that the size of $\Delta^2(k)$ is set purely by $H$ and $\epsilon$ at horizon exit, and these only weakly depend on $k$ (through their weak dependence on time) during near-exponential inflation. This is what ensures the approximate scale-invariance of the primordial power spectrum which inflation predicts for the later universe. To pin down the value of $n_s$ more precisely we take the power-law form $\Delta^2_\Phi = A k^{n_s - 1}$, for which deviations from scale invariance may be computed by evaluating

$$n_s - 1 \equiv \left.\frac{{\rm d} \ln \Delta^2_\Phi}{{\rm d} \ln k}\right|_{\rm he} . \tag{2.51}$$

To evaluate this during slow-roll inflation use the condition $k = aH$ (and the approximate constancy of $H$ during inflation) to write ${\rm d} \ln k = H {\rm d}t$. Since the right-hand side of eq. (2.48) depends on $k$ and $t$ only through its dependence on $\varphi$, it is

[14] Of course, $V$ can be much smaller if $\epsilon$ is smaller as well, or if primordial fluctuations actually come from another source.

convenient to use the slow-roll equations, eq. (1.70) to further change variables from $t$ to $\varphi$: $\mathrm{d}t = \mathrm{d}\varphi/\dot{\varphi} \simeq -(3H/V')\,\mathrm{d}\varphi$, and so

$$\frac{\mathrm{d}}{\mathrm{d}\ln k} = -M_p^2\left(\frac{V'}{V}\right)\frac{\mathrm{d}}{\mathrm{d}\varphi} = \sqrt{2\epsilon}\,M_p\,\frac{\mathrm{d}}{\mathrm{d}\varphi}\,. \tag{2.52}$$

Performing the $\varphi$ derivative using (2.48) finally gives the following relation between $n_s$ and the slow-roll parameters, $\epsilon$ and $\eta$

$$n_s - 1 = -6\epsilon + 2\eta\,, \tag{2.53}$$

where the right-hand side is evaluated at $\varphi = \varphi_{\rm he}$. For single-field models the right-hand side is negative and typically of order 0.01, agreeing well with the measured value $n_s \simeq 0.96$.

### 2.2.6 Tensor fluctuations

A similar story goes through for the tensor fluctuations, though without the complications involving mixing between $\delta\chi$ and $\Phi$. Tensor modes are also directly generated by quantum fluctuations, in this case where the vacuum is the quantum state of the graviton part of the Hilbert space. Although tensor fluctuations have not yet been observed, they are potentially observable through the polarization effects they produce as CMB photons propagate through them to us from the surface of last scattering.

Just like for scalar fluctuations, for each propagating mode the amplitude of fluctuations in the field $h_{ij}$ is set by $H/(2\pi)$, but because there is no longer a requirement to mix with any other field (like $\Phi$), the power spectrum for tensor perturbations depends only on $H^2$ rather than $H^2/\epsilon$. Repeating the above arguments leads to the following dimensionless tensor power spectrum

$$\Delta_h^2(k) = \frac{8}{M_p^2}\left(\frac{H}{2\pi}\right)^2 = \frac{2V}{3\pi^2 M_p^4}\,. \tag{2.54}$$

Should both scalar and tensor modes be measured, a comparison of their amplitudes provides a direct measure of the slow-roll parameter $\epsilon$. This is conventionally quantified in terms of a parameter $r$, defined as a ratio of the scalar and tensor power spectra

$$r := \frac{\Delta_h^2}{\Delta_\Phi^2} = 16\,\epsilon\,. \tag{2.55}$$

The absence of evidence for these perturbations to date places a relatively weak upper limit: $r \lesssim 0.10$, and so $\epsilon < 0.007$.

The detection of tensor modes in principle also allows a measurement of the $k$ dependence of their power spectrum. This is usually quantified in terms of a tensor spectral index, $n_T$, defined by

$$n_T \equiv \frac{\mathrm{d}\ln\Delta_h^2}{\mathrm{d}\ln k} = -2\epsilon = -\frac{r}{8}\,, \tag{2.56}$$

where the second-last equality evaluates the derivative within inflation as before by changing variables from $k$ to $\varphi$. This result is understood to be evaluated at the epoch when observable modes leave the horizon during inflation, $\varphi = \varphi_{\rm he}$.

Ultimately single-field models have three parameters: $\epsilon$, $\eta$ and the Hubble scale during inflation, $H_I$. But the scalar and tensor fluctuation spectra provide four observables: $A_s$, $A_T$, $n_s$ and $n_T$. The ability to describe these four observables in terms of three parameters implies that the relation $n_T = -r/8$ given in (2.56) is a robust prediction shared by all single-field slow-roll inflationary models.

# 3 EFT issues

It may not yet be clear how EFT methods enter into the beautiful story presented above, but this section argues EFT methods are actually used throughout (as is also typically true essentially everywhere else in physics). Since these lectures are being delivered in a school entirely devoted to EFTs the logic of this section is not to explain what an EFT is (such as they arise in areas like chiral perturbation theory), but rather to sketch some of the issues that come up when they are applied to gravity- and cosmology-specific problems.

In my opinion the lesson of these applications is twofold. First, there is no evidence (yet) for 'gravitational exceptionalism:' the idea that gravity is fundamentally different from all other interactions, and so there is nothing to learn from experience in other settings. The second lesson is that EFT applications to gravity can sometimes more resemble effective descriptions of particles moving through a medium than they do the traditional uses of EFTs in particle physics. As such they can be mind-broadening to those of us who approach the subject with a particle-physics training.

Each of the subsections addresses different kinds of examples of this, in turn.

## 3.1 General relativity as an EFT

The most important use of EFT methods in gravity-related problems is the one described in this subsection: the justification of the semiclassical approximation that

underpins almost all theoretical approaches. Although we usually think of gravitational interactions as being classical, a question less often asked is why it should be (and, if so, what is the small parameter that suppresses quantum effects).

The claim made here is that the issues for gravitational systems in may ways resemble those arising in nonlinear sigma-models,

$$\mathcal{L} = -\frac{f^2}{2}\; G_{ij}(\phi)\, \partial_\mu \phi^i\, \partial^\mu \phi^j \,, \tag{3.1}$$

such as describe Goldstone (and pseudo-Goldstone) bosons (including those studied in chiral perturbation theory). This similarity arises because both are non-renormalizable, in that their interactions involve inverse powers of a mass scale ($f$ for the sigma-model and $M_p$ for gravity) and both are dominated at low energies by interactions involving only two derivatives but many powers of the interacting fields.

Both of these properties lose their power to paralyze once it is recognized that the action should really also include all possible kinds of higher-derivative interactions, and it is recognized that predictive power is only possible for low-energy observables relative to $f$ (or $M_p$). For gravity this leads one to regard General Relativity as the leading part of what might be called (in analogy to the Standard Model Effective Field Theory – or SMEFT) the General Relativity Effective Field Theory – or GREFT.

### 3.1.1 GREFT

To see how this works in detail for gravity we apply to General Relativity the same steps seen in your other classes for sigma models.

The low-energy degrees of freedom in this case are gravitons, whose field is the metric, $g_{\mu\nu}$, of spacetime itself. The low-energy symmetries that constrain the form of the action are general covariance and local Lorentz invariance. Invariance under these symmetries dictate the metric can appear in the action only through curvature invariants built from the Riemann tensor and its contractions and covariant derivatives. The Riemann curvature tensor is defined by

$$\begin{aligned} R^\mu{}_{\nu\rho\lambda} &= \partial_\lambda \Gamma^\mu_{\nu\rho} + \Gamma^\mu_{\lambda\alpha}\Gamma^\alpha_{\nu\rho} - (\lambda \leftrightarrow \rho) \\ \text{with} \quad \Gamma^\mu_{\nu\lambda} &= \frac{1}{2}\, g^{\mu\beta}\Big(\partial_\nu g_{\beta\lambda} + \partial_\lambda g_{\beta\nu} - \partial_\beta g_{\nu\lambda}\Big)\,. \end{aligned} \tag{3.2}$$

$$\tag{3.3}$$

and its only independent contractions are the Ricci curvature tensor $R_{\mu\nu} = R^\alpha{}_{\mu\alpha\nu}$ and its trace $R = g^{\mu\nu}R_{\mu\nu}$, where the inverse metric, $g^{\mu\nu}$, satisfies $g^{\mu\nu}g_{\nu\lambda} = \delta^\mu_\lambda$. What is

important in what follows about these definitions is that, although complicated, the curvature tensors involve precisely two derivatives of the metric.

GREFT is defined (as usual) by writing down a local action involving all possible powers of derivatives of the metric, which general covariance then requires must be built from powers of the curvature tensors and their derivatives. This leads to the following effective lagrangian:

$$-\frac{\mathcal{L}_{\rm GREFT}}{\sqrt{-g}} = \lambda + \frac{M_p^2}{2}\,R + c_{41}\,R_{\mu\nu}\,R^{\mu\nu} + c_{42}\,R^2 + c_{43}\,R_{\mu\nu\lambda\rho}R^{\mu\nu\lambda\rho} + c_{44}\,\Box R + \frac{c_{61}}{m^2}\,R^3 + \frac{c_{62}}{m^2}\partial_\mu R\,\partial^\mu R + \cdots\,, \tag{3.4}$$

where $\sqrt{-g} = \sqrt{-\det g_{\mu\nu}}$, as usual. The first line here includes all possible terms involving two or fewer derivatives, and is the Einstein-Hilbert action of General Relativity, with cosmological constant $\lambda$. The second line includes all possible terms involving precisely four derivatives, and (for brevity) the third line includes only the first two representative examples of the many possible terms involving six or more derivatives.

The first, cosmological constant, term in eq. (3.4) is the only one with no derivatives. Its appearance complicates power-counting arguments (in much the same was as does the appearance of a scalar potential when power-counting with a sigma-model). They cause problems if their coefficients are similar in size as for the two-derivative terms, and the puzzle of why this should be true in Nature is a well-known problem [15]. For simplicity of presentation the cosmological constant term is simply dropped in the power-counting argument that follows. Once this is done the leading term in the derivative expansion is the Einstein-Hilbert term of General Relativity. Its coefficient defines Newton's constant (and so also the Planck mass, $M_p^{-2} = 8\pi G$).

The constants $c_{dn}$ are dimensionless couplings, with the convention that $d$ counts the number of derivatives of the corresponding effective operator and $n = 1, \cdots, N_d$ runs over the number of such couplings. These couplings are dimensionless because the explicit mass scales, $m$ and $M_p$, are extracted to ensure this is so. Often one sees this action written with only the Planck scale appearing, *i.e.* with $m = M_p$. However, as is usual in an EFT, the scale $m$ is usually of order the lightest particle integrated out to produce this effective theory, leaving only the metric as the variable. Since it is the *smallest* such a mass that dominates, $m$ is generically expected to be much smaller than $M_p$. (For applications to the solar system $m$ might be the electron mass; for applications to pose-nucleosynthesis Big-Bang cosmology $m$ might be of order the QCD scale, and so on.) Of course, contributions like $m^2 R$ or $R^3/M_p^2$ could also exist, but

these are completely negligible compared to the terms displayed in eq. (3.4). The central point of EFT methods is that the consequences of (3.4) should be explored as low-energy expansion in powers of $q/m$ and $q/M_p$, where $q$ is a typical energy/momentum characterizing the observables of interest.

**Redundant interactions**

Just as is true in SMEFT, to save needless effort one should eliminate those redundant interactions that can be removed by integrating by parts or performing a field redefinition. As discussed in your other lectures (see also [2]), the freedom to perform field redefinitions allows the dropping of any terms that vanish when evaluated at solutions to the lowest-order equations of motion. The freedom to drop total derivatives allows us to set the coupling $c_{44}$ to zero, as well (in 4 dimensions) as $c_{43}$. (For $c_{44}$ this can be done because $\sqrt{-g}\,\Box R$ is a total derivative, and for $c_{43}$ the relevant observation is that the quantity

$$\sqrt{-g}\; X = \sqrt{-g}\Big(R_{\mu\nu\lambda\rho}R^{\mu\nu\lambda\rho} - 4R_{\mu\nu}R^{\mu\nu} + R^2\Big)\,, \tag{3.5}$$

integrates to give a topological invariant in 4 dimensions, and so is locally also a total derivative. It is therefore always possible to replace, for example, $R_{\mu\nu\lambda\rho}R^{\mu\nu\lambda\rho}$ in the 4-derivative effective lagrangian with the linear combination $4\,R_{\mu\nu}R^{\mu\nu} - R^2$, with no consequences for any observables, provided these observables are insensitive to the overall topology of spacetime (such as are the classical equations, or perturbative particle interactions).

The freedom to perform field redefinitions also allows the removal of the other two 4-derivative terms. This is because (in the absence of other, matter, fields) the lowest order equations of motion are $R_{\mu\nu} = 0$, and the remaining terms vanish when this is imposed. For pure gravity (without a cosmological constant) the first nontrivial effective interaction involves more than 4 derivatives, such as the term proportional to the cube of the Riemann tensor. This irrelevance of all of the 4-derivative terms must be re-examined once matter fields are included, however, since once these are included $R_{\mu\nu}$ need no longer vanish.

### 3.1.2 Power Counting

In any EFT the central question asks which interactions are relevant when computing observables at a specific order in the low-energy expansion that controls the low-energy expansion in powers of $q/m$ and $q/M_p$. Because of the similarity in the structure of derivatives appearing in sigma models and General Relativity, power-counting for the

two types of theories is very similar. This section briefly recaps the result without repeating the details (see however [2]), highlighting those features that differ.

To this end start by considering the interactions of gravitons propagating in flat space (returning to curved space below). In this case we expand $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$ and identify propagators and interactions for perturbative calculations in the usual way. For the purposes of this power counting all we need to know about the curvatures is that they each involve all possible powers of $h_{\mu\nu}$, but with only precisely two derivatives. Consider an arbitrary graph that contributes at $L$ loops to the $E$-point graviton-scattering amplitude, $\mathcal{A}_E(q)$, performed with energy $q$. Suppose also the graph contains $V_{id}$ vertices involving $d$ derivatives and the emission or absorption of $i$ gravitons. Using arguments identical to those used for sigma models in your other lectures leads to the following dependence[15] of $\mathcal{A}_E(q)$ on the scales $q$, $m$ and $M_p$:

$$\mathcal{A}_E(q) \sim q^2 M_p^2 \left(\frac{1}{M_p}\right)^E \left(\frac{q}{4\pi M_p}\right)^{2L} \prod_i \prod_{d>2} \left[\frac{q^2}{M_p^2}\left(\frac{q}{m}\right)^{(d-4)}\right]^{V_{id}} . \tag{3.6}$$

Notice that since $d$ is even for all of the interactions, the condition $d > 2$ in the product implies there are no negative powers of $q$ in this expression.

Eq. (3.6) shows that the weakness of a graviton's coupling (much like the weak couplings of a Goldstone boson) comes purely from the low-energy approximations, $q \ll M_p$ and $q \ll m$. It is also clear that even though the ratio $q/m$ could be much larger than $q/M_p$, it only arises in $\mathcal{A}_E$ once contributions from at least curvature-cubed interactions are included (for which $d = 6$).

Furthermore (3.6) shows that the dominant contributions to low-energy graviton scattering amplitudes correspond to graphs with $L = 0$ and $V_{id} = 0$ for all $d > 2$. That is to say, graphs built using ony tree graphs constructed purely from the Einstein-Hilbert ($d = 2$) action: it is classical General Relativity that governs the low-energy dynamics of gravitational waves.

But EFTs excel when computing next-to-leading contributions. In this case these come in one of the following two ways. Either:

- $L = 1$ and $V_{id} = 0$ for any $d \neq 2$ but $V_{i2}$ is arbitrary, or
- $L = 0$, $\sum_i V_{i4} = 1$, $V_{i2}$ is arbitrary, and all other $V_{id}$ vanish.

[15] Technical point: as is usually the case this power counting result is computed in dimensional regularization, since not including a spurious cutoff scale makes arguments based on dimensional analysis particularly simple.

That is, the next to leading contribution is found using one-loop graphs using only the interactions of General Relativity, or by working to tree level and including precisely one insertion of a curvature-squared interaction in addition to any number of interactions from GR. Both of these are suppressed compared to the leading term by a factor of $(q/M_p)^2$. The next-to-leading tree graphs provide precisely the counter-terms required to absorb the UV divergences in the one-loop graphs. And so on.

What this shows is that the small parameter that controls the loop expansion (*i.e.* the semi-classical expansion) for graviton scattering is the ratio $q^2/(4\pi M_p)^2$; the semiclassical approximation *is* the low-energy approximation.

But the above argument was made specifically for gravitons propagating in flat space. How reliable should these power-counting arguments be for drawing conclusions for more general curved environments? Related to this, how important is it to be able to work in momentum space, as is usually done in sigma-model type arguments (and those adapted from them to gravity)?

The issue of momentum space can be put aside, because the arguments for sigma models can equally well be made in position space. The key estimate made to arrive at (3.6) is based on dimensional analysis: all of the factors of $m$ and $M_p$ are tracked by counting how they appear as factors in propagators and vertices, and the remaining dimensions are all filled in as the common low-energy scale $q$. The analogous argument works also in position space, provided there is also only one scale $q$ that characterizes the observables of interest in the low-energy theory.[16]

Physically, the equivalence of the short-distance position-space and high-energy momentum-space estimates happens because the high-energy contributions arise due to the propagation of modes having very small wavelength, $\lambda$. Provided this wavelength is very small compared with the local radius of curvature, $r_c$, particle propagation behaves just as if it had taken place in flat space. One expects the most singular behaviour to be just as for flat space, with curvature effects appearing in subdominant corrections as powers of $\lambda/r_c$.

It is often true that the low-energy gravitational system is characterized by a single scale. For cosmological models this scale is often the Hubble scale $q \sim H$. (For black holes it is instead $q \sim r_s^{-1}$ where $r_s = 2GM = M/(4\pi M_p^2)$ is the Schwarzschild radius.) In this case the above power-counting arguments imply the semiclassical expansion

[16] General EFT arguments still apply when there is more than one scale, but are more complicated. Indeed much of the complications encountered in other lectures when non-relativistic particles are present can be traced to their having more than a single scale, and the same is true for non-relativistic particles interacting with gravity.

arises as powers of $H^2/(4\pi M_p)^2$ (or $(4\pi M_p r_s)^{-2} \sim (M_p/M)^2$ in the case of black holes). We require $H/M_p \ll 1$ (or $M \gg M_p$ for black holes) in order to believe inferences about their properties using semiclassical methods.

## 3.2 Cosmology-specific issues

Besides issues specific to gravity, use of EFTs in cosmology can also involve other complications that are often not seen in particle physics.

### 3.2.1 EFTs with time-dependent backgrounds

An issue specific to cosmology arises due to the appearance there of time-dependent backgrounds. The issue asks: if EFTs are defined by dividing systems into low- and high-energy states how can they be defined in time-dependent problems where energy is not conserved? The short version of this section is that time-dependence (in gravity as well as elsewhere) time-dependence always imposes additional restrictions on the domain of validity of EFTs, most important of which is usually the requirement that the background time-dependence should be adiabatic. (That is, $\dot{\phi}/\phi$ should be smaller than the UV scales of interest, for every time-dependent field $\phi$ in the problem.)

Adiabatic motion is important because in this case the existence of an approximately conserved $H(t)$ for any given time defines both an approximate ground state and an energy in terms of which the low-energy/high-energy split can be defined. Once the system is partitioned in this way into low-energy and high-energy state, one can ask whether a purely low-energy description of time evolution is possible using only a low-energy, local effective lagrangian. The main danger is that the time evolution of the system need not keep low-energy states at low energies, or high-energy states at high energies. For instance, this could happen if the background's time-dependence is rapid enough to allow particle-production of what were regarded as high-energy states. Or it could happen that the gap between high and low energies decreases with time, such as if there is high- and low-energy states were to cross one another as time evolves.

A related issue can arise if there is a transfer of states from high-energy to low-energy as the dividing line between them, $\Lambda(t)$, evolves. For example, this could happen for a charged particle in a decreasing magnetic field if the effective theory is set up so that the dividing energy, $\Lambda(t)$, between low- and high-energies is not similarly time dependent. In this case then Landau levels continuously enter the low-energy theory as the magnetic field strength wanes. Such a migration of states can also happen in cosmology, such as during an inflationary phase (the so-called trans-Planckian 'problem'). This usually is only a problem for the effective-theory formulation if the states which

enter in this way are not in their adiabatic ground state when they do so. If they are in their adiabatic ground state they do not affect low-energy observables, but if they are not they can since then new physical excitations appear at low energies.

What emerges from this is that EFTs can make sense despite the presence of time-dependent backgrounds, provided one can focus on the evolution of low-energy states, ($q < \Lambda(t)$), without worrying about losing probability into high-energy states ($q > \Lambda(t)$). This can often be ensured if the background time evolution is sufficiently adiabatic.

### 3.2.2 Predicting background evolution with EFTs

There is another issue at stake when using EFTs in cosmology (or other time-dependent settings). Up to now the evolution of the background field is regarded as being given, and the EFT issues of the previous section are to do with understanding how to split the system into low and high energy states relative to an adiabatic energy defined in the presence of this time-dependent background.

But it is often also of interest to know how the background itself responds to events within time-dependent systems. For instance the background might back-react in response to changes in the state of fluctuations with which it interacts. This can also be amenable to EFT analysis, often by solving self-consistently for the background using the field equations of the low-energy theory. Central to this approach is the assumption that solutions to field equations within an EFT actually capture the behaviour of solutions to field equations within the full theory.

Need this always be true? This section argues in general the answer is 'no', although it usually is for adiabatic motion.

To see why EFTs and UV completions can agree on their solutions to the equations of motion one must hark back to the definitions of the EFT itself. (The EFT formulation used here follows the review [16].) Consider therefore a theory with high-energy and low-energy fields $h$ and $\ell$, with action $S(h, \ell)$. We wish to integrate out $h$ to obtain the effective action, $S_{\rm eff}(\ell)$, to examine its equations of motion. For simplicity we do so at the classical level, in which case integrating out $h$ is equivalent to solving its classical field equations as a function of the light field, $h_c(\ell)$ and plugging the result back into the original action:

$$S_{\rm eff}[\ell] = S[h_c(\ell), \ell]\,, \qquad \text{where} \quad \left(\frac{\delta S}{\delta h}\right)_{h=h_c(\ell)} = 0\,. \tag{3.7}$$

(Exercise: verify this statement explicitly by showing that it is equivalent to integrating out $h$ using only tree graphs.)

An immediate consequence of the above derivation seems to be that any solution to the low-energy EFT

$$\left(\frac{\delta S_{EFT}}{\delta \ell}\right)_{\ell=\ell_c} = 0\,, \tag{3.8}$$

must also be extrema of the full theory, by virtue of the choice $h = h_c(\ell)$. How can this argument ever fail?

The key step in deriving any EFT, glossed over in the previous paragraphs, is the necessity of expanding to some finite order in powers of the heavy mass scale, $1/M$. It is only after this expansion that an effective action like (3.7) is given by a local lagrangian density. Because of this we should only trust the equations of motion of any local EFT up to the same order in powers of $1/M$. Solutions of the full theory can differ from those of the effective theory if they are not captured by such a $1/M$ expansion.

### 3.2.3 Exorcising the ghosts

It is actually a good thing that the solutions to an EFT are not completely equivalent to solutions to the full theory from which the EFT is derived. One upside is that EFTs often involve higher time derivatives, and so naively should generically have unstable runaway solutions, even if the underlying theory has no such instabilities.

To see why instabilities might arise within the EFT consider the following toy effective lagrangian:

$$\frac{L}{v^2} = \frac{1}{2}\,\dot\theta^2 + \frac{1}{2M^2}\,\ddot\theta^2\,, \tag{3.9}$$

whose variation $\delta L = 0$ gives the linear equation of motion

$$-\ddot\theta + \frac{1}{M^2}\,\ddddot\theta = 0\,. \tag{3.10}$$

The general solution to this equation is

$$\theta = A + Bt + Ce^{Mt} + De^{-Mt}\,, \tag{3.11}$$

where $A$, $B$, $C$ and $D$ are integration constants.

Now comes the main point. Only the solutions with $C = D = 0$ go over to the solutions to the lowest-order field equation, obtained from the $M \to \infty$ lagrangian, $L_0 = \frac{1}{2}\dot\theta^2$. The others make no sense at any finite order of $1/M$ because for them the $\dot\theta^2$ and $\ddot\theta^2$ terms are always comparably large. Since a local EFT is only meant to capture the full theory order-by-order in $1/M$ these exponential solutions should not be expected to be reproducing the low-energy approximation of the full theory.

### 3.2.4 Open systems

EFTs applied to gravitational systems can surprise in other ways as well. In particular, during inflation we have seen that the main observational consequences are tied up with super-Hubble modes, for which $k/a \ll H$. Since these are the longest-wavelength modes in the system the effective action that describes them has long been sought as the most efficient way to capture inflationary predictions in as model-independent way as possible. But no such an effective action was ever found.

This doesn't mean an effective description does not exist, it just turns out not to be described by an effective action [17]. This unusual situation arises because during inflation the long-wavelength modes are an *open* system, in that modes are continually moving from sub-Hubble to super-Hubble throughout the inflationary epoch. This should be contrasted with the usual situation with a Wilsonian effective theory, for which high- and low-energy states are forbidden from transitioning into one another by energy conservation.

Because of this mode migration the long- and short-distance sectors can interact in more complicated ways than are normally entertained, such as by entangling and/or decohering with one another. The appropriate language for describing long-wavelength modes in this kind of situation is to use the reduced density matrix, $\varrho_L = \mathrm{Tr}_S\, \rho$, in which the full system's density matrix is traced over the unwatched (in this case, short-wavelength) sector. It turns out that $\varrho_L$ evolves in time according to what is called a Lindblad equation, which need not be writable as a Liouville equation for some choice of effective Hamiltonian.

Using these kinds of arguments it is possible to show that the leading effective description of fluctuations amongst super-Hubble modes during inflation is given by a Fokker-Planck equation, the description of which is called 'stochastic inflation'. Although their description goes beyond the scope of these lectures, the evidence now is that stochastic inflation does a best job capturing the late-time evolution of super-Hubble modes.

## 3.3 EFT of inflationary fluctuations

There is yet another kind of effective description that arises within the inflationary literature, one that has come to be known as 'the' Effective Theory of Inflation. These notes close with a very brief summary of this specific theory, in order to put it into its context in the broad EFT pantheon.

The EFT of Inflation is aimed at single-field inflationary models including, but not restricted to, the simple models considered above. The starting point of this theory is

the observation that when the single field rolls homogeneously, $\varphi(t)$, its rolling (and the geometry the energy in this rolling creates) provides the clock that breaks the time-translation invariance the flat spacetime otherwise would have had in the absence of $\varphi$. In many ways $\varphi$ acts as the Goldstone boson for this breaking, though it is ultimately eaten by the metric, which is the gauge field for local translations.

TO BE CONTINUED....

## References


[1] C. P. Burgess, "Lectures on Cosmic Inflation and its Potential Stringy Realizations," Class. Quant. Grav. **24** (2007) S795 [PoS P **2GC** (2006) 008] [PoS CARGESE **2007** (2007) 003] doi:10.1088/0264-9381/24/21/S04 [arXiv:0708.2865 [hep-th]].

[2] C. P. Burgess, "Quantum gravity in everyday life: General relativity as an effective field theory," Living Rev. Rel. **7** (2004) 5 doi:10.12942/lrr-2004-5 [gr-qc/0311082].

[3] C.P. Burgess and M. Williams, "Who You Gonna Call? Runaway Ghosts, Higher Derivatives and Time-Dependence in EFTs," XXXX [arXiv:1404.2236 [gr-qc]].

[4] S. Weinberg, *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity*, Wiley 1972.

[5] C. W. Misner, J. A. Wheeler and K. S. Thorne, *Gravitation*, W. H. Freeman & Company 1973.

[6] D. Baumann, "Inflation," doi:10.1142/9789814327183_0010 arXiv:0907.5424 [hep-th].

[7] F. L. Bezrukov and M. Shaposhnikov, "The Standard Model Higgs boson as the inflaton," Phys. Lett. B **659** (2008) 703 doi:10.1016/j.physletb.2007.11.072 [arXiv:0710.3755 [hep-th]].

[8] C. P. Burgess, H. M. Lee and M. Trott, "Power-counting and the Validity of the Classical Approximation During Inflation," JHEP **0909** (2009) 103 doi:10.1088/1126-6708/2009/09/103 [arXiv:0902.4465 [hep-ph]];
J. L. F. Barbon and J. R. Espinosa, "On the Naturalness of Higgs Inflation," Phys. Rev. D **79** (2009) 081302 doi:10.1103/PhysRevD.79.081302 [arXiv:0903.0355 [hep-ph]];
C. P. Burgess, H. M. Lee and M. Trott, "Comment on Higgs Inflation and Naturalness," JHEP **1007** (2010) 007 doi:10.1007/JHEP07(2010)007 [arXiv:1002.2730 [hep-ph]].

[9] V. Mukhanov, *Physical Foundations of Cosmology*, Cambridge University Press (2005).

[10] XXX CITE HERE REFERENCES PROVING CONSTANT $\zeta$

[11] A. H. Guth and S. Y. Pi, "The Quantum Mechanics Of The Scalar Field In The New Inflationary Universe," Phys. Rev. D **32**, 1899 (1985);
M. A. Sakagami, "Evolution From Pure States Into Mixed States In De Sitter Space," Prog. Theor. Phys. **79**, 442 (1988);
L. P. Grishchuk and Y. V. Sidorov, "On The Quantum State Of Relic Gravitons," Class. Quant. Grav. **6** (1989) L161;
R. H. Brandenberger, R. Laflamme and M. Mijic, "Classical Perturbations From Decoherence Of Quantum Fluctuations In The Inflationary Universe," Mod. Phys. Lett. A **5**, 2311 (1990);
E. Calzetta and B. L. Hu, "Quantum fluctuations, decoherence of the mean field, and structure formation in the early universe," Phys. Rev. D **52**, 6770 (1995) [gr-qc/9505046];
D. Polarski and A. A. Starobinsky, "Semiclassicality and decoherence of cosmological perturbations," Class. Quant. Grav. **13**, 377 (1996) [gr-qc/9504030];
F. C. Lombardo and D. Lopez Nacir, "Decoherence during inflation: The generation of classical inhomogeneities," Phys. Rev. D **72**, 063506 (2005) [gr-qc/0506051];
C.P. Burgess, R. Holman and D. Hoover, "On the Decoherence of Primordial Fluctuations During Inflation," [astro-ph/0601646].

[12] C. Kiefer, D. Polarski and A. A. Starobinsky, "Quantum-to-classical transition for fluctuations in the early universe," Int. J. Mod. Phys. D **7**, 455 (1998) [gr-qc/9802003];
J. Lesgourgues, D. Polarski and A. A. Starobinsky, "Quantum-to-classical transition of cosmological perturbations for non-vacuum initial states," Nucl. Phys. B **497**, 479 (1997) [gr-qc/9611019].

[13] C. P. Burgess, R. Holman, G. Tasinato and M. Williams, "EFT Beyond the Horizon: Stochastic Inflation and How Primordial Quantum Fluctuations Go Classical," JHEP **1503** (2015) 090 doi:10.1007/JHEP03(2015)090 [arXiv:1408.5002 [hep-th]].
H. Collins, R. Holman and T. Vardanyan, "The quantum Fokker-Planck equation of stochastic inflation," arXiv:1706.07805 [hep-th].

[14] L. P. Grishchuk and Y. V. Sidorov, "Squeezed Quantum States Of Relic Gravitons And Primordial Density Fluctuations," Phys. Rev. D **42**, 3413 (1990);
A. Albrecht, P. Ferreira, M. Joyce and T. Prokopec, "Inflation and squeezed quantum states," Phys. Rev. D **50**, 4807 (1994) [astro-ph/9303001].

[15] See, for example:
S. Weinberg, *Rev. Mod. Phys.* **61** (1989) 1;
E. Witten, "The Cosmological constant from the viewpoint of string theory," [hep-ph/0002297];
J. Polchinski, "The Cosmological Constant and the String Landscape,"

[hep-th/0603249];
C.P. Burgess, "The Cosmological Constant Problem: Why it is Hard to Get Dark Energy from Micro-Physics," in the proceedings of the Les Houches School *Cosmology After Planck*, [arXiv:1309.4133];
T. Banks, "Supersymmetry Breaking and the Cosmological Constant," Int. J. Mod. Phys. A **29** (2014) 1430010 [arXiv:1402.0828 [hep-th]];
A. Padilla, "Lectures on the Cosmological Constant Problem," [arXiv:1502.05296 [hep-th]].

[16] C. P. Burgess, Introduction to Effective Field Theory, Ann. Rev. Nucl. Part. Sci. 57 (2007) 329 [hep-th/0701053].

[17] C.P. Burgess, R. Holman, G. Tasinato and M. Williams, "EFT Beyond the Horizon: Stochastic Inflation and How Primordial Quantum Fluctuations Go Classical," JHEP **1503** (2015) 090 [arXiv:1408.5002 [hep-th]].
C.P. Burgess, R. Holman and G. Tasinato, "Open EFTs, IR Effects & Late-Time Resummations: Systematic Corrections in Stochastic Inflation," JHEP **1503** (2015) 090 [arXiv:1512.00169 [gr-qc]].